BIOMETRIKA

BIOMETRIKA

FOUNDED BY

W. F. R. WELDON, FRANCIS GALTON AND KARL PEARSON

MANAGING EDITOR

E. S. PEARSON

ASSOCIATE EDITORS

M. G. KENDALL JOHN WISHART

in consultation with

HARALD CRAMÉR R. C. GEARY

J. B. S. HALDANE

VOLUME 37

1950

ISSUED BY

THE BIOMETRIKA OFFICE, UNIVERSITY COLLEGE, LONDON

PRINTED AT THE UNIVERSITY PRESS, CAMBRIDGE

PRINTED IN GREAT BRITAIN

CONTENTS OF VOLUME 37

Memoirs

PAGES

I. Periodogram analysis and continuous spectra. By M. S. Bartlett. With five Figures in the Text . . . . 1—16

II. Notes on continuous stochastic phenomena. By P. A. P. Moran 17—23

III. Accident proneness. By Major Greenwood . . . . . 24—29

IV. Organic correlation and allometry. By K. A. Kermack and J. B. S. Haldane . . . . 30—41

V. The probability integral transformation when the variable is discontinuous. By F. N. David and N. L. Johnson. With one Figure in the Text . . . . 42—49

VI. How balanced incomplete block designs may be made to furnish orthogonal estimates in weighing designs. By K. S. Banerjee 50—58

VII. The power of the χ^2 index of dispersion test when Neyman's contagious distribution is the alternate hypothesis. By G. I. Bateman. With one Figure in the Text . . . 59—63

VIII. Power of the modified t-test (u-test) based on range. By E. Lord . 64—77

IX. The use of mean range as an estimator of variance in statistical tests. By P. B. Patnaik . . . . 78—87

X. Some notes on the use of range. By E. S. Pearson . . . 88—92

XI. On the statistical independence of quadratic forms in normal variates. By A. C. Aitken . . . . 93—96

XII. Two combinatorial tests of whether a sample has come from a given population. By F. N. David. With one Figure in the Text . . . . 97—110

XIII. Maximum likelihood and the efficiency of the method of moments. By L. R. Shenton . . . . 111—116

XIV. Fiducial limits of the parameter of a discontinuous distribution. By W. L. Stevens. With two Figures in the Text . . . 117—129

XV. Extension of the Neyman-Pearson theory of tests to discontinuous variates. By K. D. Tocher . . . . 130—144

XVI. A simplified form of Sheppard's correction formulae. By H. O. Hartley . . . . 145—148

XVII. Some theorems in least squares. By R. L. Plackett . . . 149—157

XVIII. Weighted probits and their use. By A. N. Black . . . 158—167

XIX. Table of the probability integral of the t-distribution. By H. O. Hartley and E. S. Pearson . . . . 168—172

XX. A simple stochastic epidemic. By Norman T. J. Bailey. With two Figures in the Text . . . . 193—202

PAGES

XXI. On the Fisher-Behrens test. By G. A. Barnard . . . . 203—207

XXII. The incomplete beta function as a contour integral and a quickly converging series for its inverse. By M. E. Wise . . . . 208—218

XXIII. On the levels of significance of the incomplete beta function and the F-distributions. By Leo A. Aroian . . . . 219—223

XXIV. On the generalized second limit-theorem in the calculus of probabilities. By K. S. Rao and David G. Kendall . . . 224—230

XXV. A note on the cumulants of Kendall's S-distribution. By H. Silverstone . . . . 231—235

XXVI. The distribution of the variance ratio in random samples of any size drawn from non-normal universes. By A. K. Gayen. With four Figures in the Text . . . . 236—255

XXVII. The comparison of percentages in matched samples. By W. G. Cochran . . . . 256—266

XXVIII. The exact partition of χ^2 and its application to the problem of the pooling of small expectations. By H. O. Lancaster . . . 267—270

XXIX. The use of range in analysis of variance. By H. O. Hartley. With one Figure in the Text . . . . 271—280

XXX. On the comparison of estimators. By N. L. Johnson. With four Figures in the Text . . . . 281—287

XXXI. A rapid method for ascertaining serial lag correlation. By Gordon D. Gibson. With seven Figures in the Text . . . . . 288—307

XXXII. The maximum F-ratio as a short-cut test for heterogeneity of variance. By H. O. Hartley . . . . 308—312

XXXIII. Tables of the χ^2-integral and of the cumulative Poisson distribution. By H. O. Hartley and E. S. Pearson . . . . 313—325

XXXIV. On a sequential t-test. By S. Rushton . . . . 326—333

XXXV. Properties of some tests in sequential analysis. By A. G. Baker. With three Figures in the Text . . . . 334—346

XXXVI. The unbiased estimation of heterogeneous error variances. By A. S. C. Ehrenberg . . . . 347—357

XXXVII. Sampling theory of the negative binomial and logarithmic series distributions. By F. J. Anscombe. With one Figure in the Text . . . . 358—382

XXXVIII. On questions raised by the combination of tests based on discontinuous distributions. By E. S. Pearson. With two Figures in the Text . . . . 383—398

XXXIX. Significance of difference between the means of two non-normal samples. By A. K. Gayen . . . . 399—408

XL. Testing for serial correlation in least squares regression. I. By J. Durbin and G. S. Watson . . . . 409—428

XLI. Distribution of 'Student'-Fisher's t in samples from compound normal functions. By Hannes Hyrenius . . . . . 429—442

Miscellanea

PAGES

(i) A method of estimating the parameters of an autoregressive time-series. By S. G. Ghurye. With two Figures in the Text . . . . . 173—178

(ii) A test for the serial independence of residuals. By P. A. P. Moran 178—181

(iii) Efficiencies of certain linear systematic statistics for estimating dispersion from normal samples. By K. R. Nair . . . . . 182—183

(iv) Apparent correlation between independent series of autocorrelated observations. By Gilbert Walker . . . . 184—185

(v) On the inversion of circulant matrices. By I. J. Good . . . . 185—186

(vi) The effect of non-normality on the z-test, when used to compare the variances in two populations. By D. J. Finch . . . . . 186—189

(vii) Review of M. G. Kendall's *Rank Correlation Methods*. By F. N. David 190

(viii) Review of *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*. Edited by J. Neyman. By F. N. David 190—191

(ix) The comparison of pairs of treatments in split-plot experiments. By J. Taylor . . . . 443—444

(x) On the best unbiased quadratic estimate of the variance. By H. Nagler 444—445

(xi) The cumulants of the first n natural numbers. By A. Stuart . . 446

(xii) Note on the χ^2 smooth test. By D. A. S. Fraser . . . . . 447—448

(xiii) An alternative form of χ^2. By F. N. David . . . . . . 448—451

(xiv) On a theorem concerning the secondary subscripts of deviations in multivariate correlation using Yule's notation. By K. N. Chandler 451—452

(xv) Review of N. L. Johnson and H. Tetley's *Statistics*, Vol. I. By R. E. Beard . . . . 453—454

Vol. 37, Parts 1 and 2 June 1950

BIOMETRIKA

FOUNDED BY

W. F. R. WELDON, FRANCIS GALTON AND KARL PEARSON

MANAGING EDITOR

E. S. PEARSON

ASSOCIATE EDITORS

M. G. KENDALL JOHN WISHART

in consultation with

HARALD CRAMÉR R. C. GEARY

J. B. S. HALDANE

ISSUED BY

COLLEGE, LONDON

VOLUME 37, PARTS 1 AND 2

JUNE 1950

PERIODOGRAM ANALYSIS AND CONTINUOUS SPECTRA

BY M. S. BARTLETT, *University of Manchester*

I. INTRODUCTION

It has been realized for some time that the classical method of searching for periodicities in time series, the so-called periodogram analysis of the series, is useless in many cases. An alternative approach by way of the autoregressive or autocorrelative structure of the series was considered in a pioneering paper by Yule (1927), and more recently by other writers (see especially M. G. Kendall (1946)).

A fuller understanding of the scope and relation of these methods is obtainable from the mathematical theory of stationary random or stochastic processes developed in the last twenty years by Wiener (1930), Khintchine (1934) and others (for an account of this theory see Lévy (1948), especially chapter IV; for further reviews of the general theory of stochastic processes see also Cramér (1947), Moyal (1949)). By *stationary* is meant, roughly speaking, that the series is oscillating or fluctuating about a constant mean; more precisely, that its distributional properties in the stochastic sense do not depend on the absolute value of the time (note that a simple harmonic series is not stationary in this sense unless its initial phase is random). For many purposes in the correlation and harmonic theory of these processes it is sufficient to assume stationarity 'to the second order', that is, in addition to the mathematical expectation or stochastic average $E\{X(t)\} =$ constant $= 0$, say, for the value of the series X at time t, the autocovariance function

$$E\{X^*(t)\,X(t+\tau)\} = w(\tau), \tag{1}$$

does not depend on t and is a function only of the interval τ. ($X^*(t)$ denotes in general the complex conjugate of $X(t)$, though in the later application here of the theory it will be assumed that the series is real.) If X is standardized by dividing by its constant standard deviation σ, assumed finite, equation (1) gives the autocorrelation function $\rho(t) \equiv w(\tau)/\sigma^2$.

If it is further assumed that $X(t)$ is continuous in mean square, that is,

$$\operatorname*{Lt}_{h\to 0} E\{|X(t+h)-X(t)|^2\} = 0,$$

a condition easily seen to be equivalent to $\rho(t)$ being continuous at $\tau = 0$, then it is known (cf. Khintchine, 1934) that the necessary and sufficient condition for $\rho(\tau)$ to be a valid autocorrelation function is that it is expressible in the form

$$\rho(\tau) = \int_{-\infty}^{\infty} e^{i\tau\omega}\,dF(\omega), \tag{2}$$

where $F(\omega)$ has the form of a cumulative distribution function, and is the integrated *spectrum* of the process. The corresponding theorem for processes defined only at discrete integral values of time is (Wold, 1938)

$$\rho_\tau = \int_{-\pi}^{\pi} e^{i\tau\omega}\,dF(\omega). \tag{3}$$

Corresponding to (2), there is an equivalent relation for $X(t)$ itself, associated with its harmonic analysis (first established in full generality by Cramér (1942; cf. Slutsky, 1938; see also Lévy, 1948, § 27, ch. IV), namely,

$$X(t) = \int_{-\infty}^{\infty} e^{it\omega}\,dZ(\omega), \tag{4}$$

where the *orthogonal* stochastic process $Z(\omega)$ has the property

$$\Delta F(\omega) = E\{Z^*(\omega+\Delta\omega)\,\Delta Z(\omega)\} = E\{\Delta Z^*(\omega)\,\Delta Z(\omega)\}, \tag{5}$$

where Δ denotes as usual a first difference, and the *stochastic* Stieltjes integral in (4) is defined as the limit in mean square of appropriate stochastic sums.

Similarly, corresponding to (3), the relation is

$$X_t = \int_{-\pi}^{\pi} e^{it\omega}\,dZ(\omega). \tag{6}$$

2. THE SPECTRA OF LINEAR PROCESSES

The relevance of the above results becomes clear when it is remembered that the general distribution function in (3) may consist of a discrete or continuous distribution, or of both. Only in the first case of a discrete distribution, apart perhaps from an additional purely random error or 'noise' leading to a *uniform* continuous distributional component, is the conception of a classical harmonic series valid. In general the possibility of any kind of continuous component to the spectrum must be allowed for.

The very generality of these results becomes embarrassing when it comes to the analysis of actual series, and further assumptions about the character of the series are usually necessary in practice before much progress can be made—hence the importance of one general type of series distinct from the classical harmonic series and having a continuous spectrum. This was studied in a special case by Yule, but a much more general form, for which the term *linear* process has been used (Bartlett, 1946), is

$$X(t) = \int_{-\infty}^{t} g(t-u)\,dY(u), \tag{7}$$

or for discrete time

$$X_t = \sum_{-\infty}^{t} g_{t-u} W_u, \tag{8}$$

where the random 'disturbances' W_u or $dY(u)$ are independent and stationary. It may readily be shown that when the coefficients g_{t-u} or $g(t-u)$ are such that X_t or $X(t)$ is well defined, then the series is stationary with autocorrelation function

$$\rho_\tau = \sigma_0^2 \sum_{u=-\infty}^{\infty} g_u^* g_{u+\tau}/\sigma^2, \tag{9}$$

or

$$\rho(\tau) = \sigma_0^2 \int_{-\infty}^{\infty} g^*(u)\,g(u+\tau)\,du/\sigma^2, \tag{10}$$

where for convenience in these formulae u runs to $-\infty$ with g_u or $g(u) \equiv 0$ for $u \leqslant 0$, and where σ_0^2 is the variance of W_u or, in the continuous time case, the variance of the integrated disturbance over unit time, $\int_t^{t+1} dY(u)$. Hence if

$$h(\omega) = \begin{cases} \sum\limits_{u=-\infty}^{\infty} e^{-iu\omega} g_u, \\ \int_{-\infty}^{\infty} e^{-iu\omega} g(u)\,du, \end{cases}$$

the spectrum, obtainable from the autocorrelation function by 'inverting' the relation (3) or (4), is found to be continuous with probability density

$$f(\omega) = \sigma_0^2 h(\omega)\,h^*(\omega)/(2\pi\sigma^2). \tag{11}$$

It should be noted that any stationary process with *continuous* spectrum can mathematically be expressed in the linear form

$$X(t) = \int_{-\infty}^{\infty} g(t-u)\, dY(u), \tag{12}$$

but, in contrast with (7), $g(u)$ does not necessarily vanish for $u < 0$ (and similarly for formula (8) in the case of discrete time), and further the $dY(u)$ are uncorrelated (orthogonal) but not necessarily independent increments (see Karhunen, 1947).

As an example, consider the simple autoregressive scheme

$$X_{t+2} + aX_{t+1} + bX_t = W_{t+2} \quad (a > 0), \tag{13}$$

for which, from the solution of this difference equation for X_t,

$$g_u = \frac{\lambda^{u+1} - \mu^{u+1}}{\lambda - \mu}, \tag{14}$$

where λ and μ are the roots (assumed complex) of $x^2 + ax + b = 0$. The ratio σ_0^2/σ^2 is given from (10) or (11) when $\tau = 0$, and it will be found that (cf. Bartlett, 1948)

$$2\pi f(\omega) = \frac{(1-b)(1-a^2+b^2+2b)}{(1+b)\{1+a^2+b^2-2b+2a(1+b)\cos\omega+4b\cos^2\omega\}} \quad (-\pi \leqslant \omega \leqslant \pi), \tag{15}$$

or if, as is more usual for real series, $f(\omega) + f(-\omega) = 2f(\omega)$ is denoted by $f_+(\omega)$ defined for $\omega \geqslant 0$, the expression in (15) is $\pi f_+(\omega)$.

If it is known that a series is, say, of the simple type (13), its autoregressive constants a and b can be estimated, and hence the spectrum calculated from formulae such as (15). However, while it is possible to test approximately the goodness of fit of the corresponding calculated autocorrelation function or 'correlogram' when compared with the observed correlogram (Quenouille, 1947), it may be required to make a direct empirical study of the spectrum without any intermediate assumptions about its particular autoregressive structure.

3. STOCHASTIC PROPERTIES OF THE PERIODOGRAM

The classical analysis of a (discrete) series of n observations consists of computing the quantity

$$H_p = A_p + iB_p = \sqrt{\frac{2}{n}} \sum_{r=1}^{n} X_r e^{2\pi i p r/n} \tag{16}$$

(by computing A_p and B_p), and hence the intensity $I_p = H_p H_p^*$. The factor $\sqrt{(2/n)}$ has been inserted to make the mean value of I_p equal to $2\sigma^2$ for a completely random series.

It is easily seen that

$$I_p = \frac{2}{n} \sum_{r=1}^{n} \sum_{u=1}^{n} X_r X_u \cos\omega(r-u), \tag{17}$$

where $\omega = 2\pi p/n$, or if

$$C_s = \frac{1}{n-s} \sum_{r=1}^{n-s} X_r X_{r+s} \quad (s > 0,\ C_{-s} = C_s),$$

then

$$I_p = 2 \sum_{s=-n+1}^{n-1} \left(1 - \frac{|s|}{n}\right) C_s \cos\omega s. \tag{18}$$

For convenience it is still assumed that $E\{X\} = 0$, but since formula (16) may be used when $E\{X\} \neq 0$, the effect of this will be noted later. When $E\{X\} = 0$, the average of equation (18) gives

$$E\{I_p\} = 2\sigma^2 \sum_{s=-n+1}^{n-1} \left(1 - \frac{|s|}{n}\right) \rho_s \cos\omega s. \tag{19}$$

As n increases, this tends for most values of ω to $2\pi\sigma^2 f_+(\omega)$, where $f_+(\omega)$ is the (continuous) spectrum of X_r.

The usual periodogram argument considers what happens when X_r contains a harmonic component of frequency Ω, so that ρ_s has a component $\lambda\cos\Omega s$. Since in this case the spectrum contains a discrete component at $\omega = \Omega$, it is evident that $E\{I_p\}$ will tend to infinity at this point. This effect is well known, but is noted here for reference. The contribution may be evaluated as $O(1/n)$ in general for $\omega \neq \Omega$, but $O(n)$ at $\omega = \Omega$, and the analysis can resolve frequencies which differ by more than $O(1/n)$. (The effect of not measuring X from $E(X)$ is now apparent; it introduces a spurious component of this type at the origin $\omega = 0$.)

It has also been recognized that as the intensity I_p fluctuates about its mean value $E(I_p) = 2\sigma^2$ for a completely random series, it is advisable when the significance of the periodogram peaks are considered to allow for such fluctuations, which are known to follow the probability law

$$P\{I_p \geqslant z\} = \exp\{-z/E(I_p)\}, \tag{20}$$

when X_r is normal (see, for example, Fisher, 1929).

Once the possibility of continuous spectra, beyond the uniform 'noise' spectrum of the completely random process, is admitted, the situation is completely changed. Not only may a non-uniform continuous spectrum have peaks which may be comparable for a finite n with the value of $E\{I_p\}$ even at the frequency Ω of a discrete harmonic component, but the fluctuations in I_p possess the remarkable property, exhibited in (20) in the particular case of the uniform spectrum, of being of the same order as the mean value $E\{I_p\}$ itself, irrespective of the length of series available for analysis. This has been noted by Slutsky (1937) and by P. J. Daniell (see the discussion to the paper by Bartlett, 1946), the second writer giving an explicit derivation which includes the property (also noted by Slutsky) that the covariance or correlation of I_p and I_q for $p \neq q$ is in general of $O(1/n^2)$. To demonstrate further the *approximate*† validity of (20) for normal X_r (or more generally for approximately normal A_p and B_p), it remains to show that the two components have equal variance and zero correlation, at least as $n \to \infty$. To do this, consider $E\{H_p^2\}$. This is

$$\begin{aligned}
E\left\{\frac{2}{n}\sum_{r=1}^{n}\sum_{n=1}^{n} X_r X_u e^{i\omega(r+u)}\right\} &= \frac{2\sigma^2}{n}\sum_{s=-n+1}^{n-1} e^{|s|i\omega}\rho_s \sum_{v=1}^{n-|s|} e^{2iv\omega} \\
&= \frac{2\sigma^2}{n}\sum_{s=-n+1}^{n-1}\rho_s \frac{e^{(|s|+2)i\omega}(1-e^{2(n-|s|)i\omega})}{1-e^{2i\omega}} \\
&= \frac{2\sigma^2}{n}\sum_{s=-n+1}^{n-1}\rho_s \frac{e^{2i\omega}(e^{|s|i\omega}-e^{-|s|i\omega})}{1-e^{2i\omega}} \qquad \text{(for } p \text{ integral)} \\
&= -\frac{2\sigma^2}{n}\sum_{s=-n+1}^{n-1}\rho_s \frac{e^{i\omega}\sin|s|\omega}{\sin\omega}.
\end{aligned} \tag{21}$$

This is exactly zero for $\rho_s = 0$ $(s \neq 0)$; but it is also $O(1/n)$ otherwise, provided that the sum $\Sigma\rho_s\sin|s|\omega$ converges as $n \to \infty$. Since

$$E\{H_p^2\} \equiv E\{A_p^2 - B_p^2\} + 2iE\{A_p B_p\},$$

the required results are thus established simultaneously.

4. SMOOTHING THE PERIODOGRAM

These results stress the futility of appraising the significance of peaks in I_p as p varies by the classical method when the spectrum is continuous, but they do not indicate how the latter is to be estimated. In such a case the standard periodogram will exhibit a wildly fluctuating appearance, a phenomenon which has often been observed in practice. Averaging over neighbouring values of p or ω was suggested by P. J. Daniell, a device which might be especially useful in automatic electrical or optical analysis of a continuous record (for simplicity the theoretical fluctuation formulae have been discussed for a discrete series of n observations, but similar formulae apply to the continuous time case). For arithmetical analysis the following alternative procedure has been suggested (Bartlett, 1948).

---

† It seems implied by Slutsky (1937, p. 133), without proof, that (20) still holds exactly, but this is clearly not so.

An average taken from m *independent* series 'lengths', if available, would possess the usual sampling property of its fluctuation being proportional to $1/\sqrt{m}$. If the lengths were taken consecutively as contiguous portions of one length of series, the correlation between quantities like H_p derived from different portions will be negligible as the length n of each portion increases, provided ρ_s decreases fast enough as s increases. For example, even between adjacent portions,

$$E\{H_p(H_p^*)'\} = 2\sigma^2 \sum_{s=-n+1}^{n-1} \left(1-\frac{|s|}{n}\right) e^{is\omega}\rho_{s+n},$$

where H and H' are obtained from the two portions, and this expression tends to zero as n increases provided ρ_s contains an exponential damping factor. Now averaging I_p for the m subseries gives the formula

$$\bar{I}_p = 2 \sum_{s=-n+1}^{n-1} \left(1-\frac{|s|}{n}\right) \bar{C}_s \cos \omega s,$$

where

$$\bar{C}_s = \frac{1}{m(n-s)} \sum_{u=0}^{m-1} \sum_{r=1}^{n-s} X_{r+nu} X_{r+s+nu} \quad (s \geqslant 0).$$

This formula ignores, however, the information available in the interval s at the end of each subseries; if this is replaced by use of the covariance estimate from the whole data, the amended formula is

$$I_p(\text{smoothed}) = 2 \sum_{s=-n+1}^{n-1} \left(1-\frac{|s|}{n}\right) C_s \cos \omega s, \tag{22}$$

where C_s is obtained from the entire series. The choice of n in this formulae (for given nm) is a compromise between reducing the fluctuations by the approximate factor $1/\sqrt{m}$ and not reducing too much the resolving power, which depends on the value of n. Formula (22) is related to the harmonic analysis of the correlogram (cf. D. G. Kendall, 1948), which would often have been calculated already, and it is sometimes convenient to use the sample autocorrelation R_s in place of C_s. In either case it is usually advisable to eliminate the spurious component at $\omega = 0$ by measuring X from the series mean.

5. EMPIRICAL VERIFICATION

Empirical verification of the validity of the suggested procedure was carried out on M. G. Kendall's artificial series (loc. cit.), for which the theoretical spectra can be calculated by means of formula (15). Some preliminary calculations reported earlier (Bartlett, 1948) also tested a cruder form of formula (22) in which the factor $1-|s|/n$ was dropped. This was found, however, to lead to relatively small but systematic oscillations which are best avoided. It was also found that the differences between working with correlations and covariances appeared negligible, and it was preferred to work with the covariances (using for this purpose sums of squares and products kindly supplied by M. G. Kendall). Since the artificial series were already measured from the mean, no correction for the mean was necessary.

Series I and III were selected as representative cases, and after testing out formula (22) on the whole series further calculations on subseries were made to check the order of both systematic and random fluctuations of the smoothed periodogram from the true spectrum. For reference the values of a and b for the two series are:

Series I (total length 480 terms): $a = -1{\cdot}1, \quad b = 0{\cdot}5.$

Series III (total length 240 terms): $a = -1{\cdot}1, \quad b = 0{\cdot}8.$

The distribution of the random disturbance W for these series was rectangular, not normal, but this does not affect the *asymptotic* distributional theory of H_p as n increases.

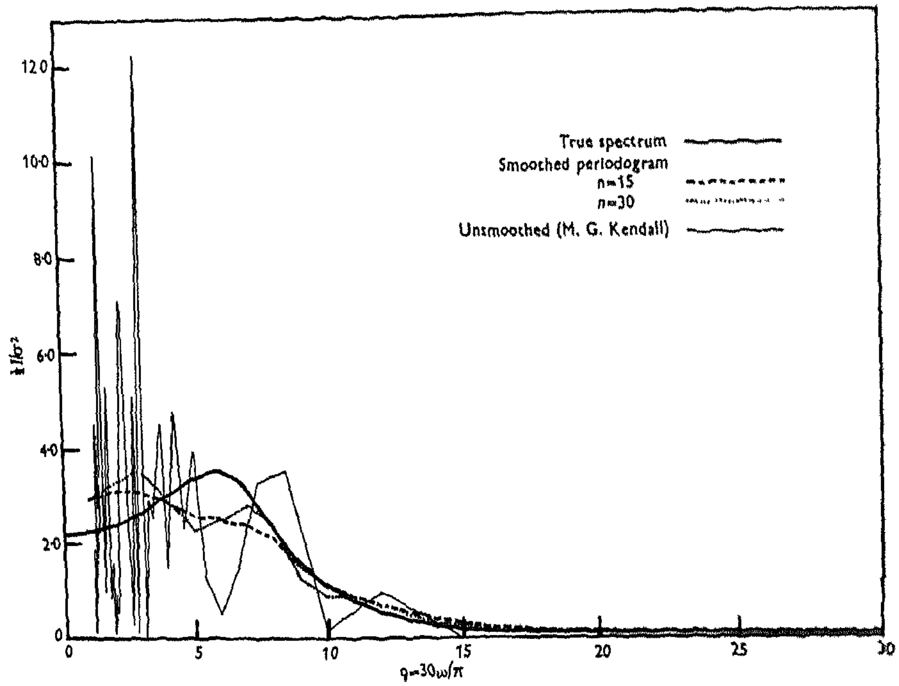

Fig. 1. Periodogram analysis of Series I—smoothed periodograms ($n = 15$ and 30) and unsmoothed (M. G. Kendall), compared with true spectrum.

In formula (22) the choice of $n = 15$, 30 and 45 for Series I gave $m = 32$, 16 and $12\frac{2}{3}$ respectively when the total length was used; and $n = 15$ and 30 for Series III gave $m = 16$ and 8. Similar calculations on subseries lengths of 60 terms for $n = 15$ and 30 corresponded of course to smaller m and hence larger fluctuations—$m = 4$ and 2 respectively for both series (these subseries calculations were confined to the more interesting half of the frequency range containing the peak of the spectrum).

The results of these computations are given in Appendix A (Tables 5–10). The intensity I was suitably standardized in the case of the covariance calculations by division by the value of the variance σ^2, and hence was comparable to $2\pi f_+(\omega)$, also shown in Tables 5 and 6. The values were calculated for $q \equiv 30\omega/\pi$ running from 1 to 30 (the corresponding values of p are $\frac{1}{2}n\omega/\pi = nq/60$). The results for the total series are also illustrated in Figs. 1 and 2, which show the smoothed periodograms for Series I and III respectively ($n = 15$ and 30) together with the true spectrum. In the case of Series I some values of the unsmoothed periodogram (suitably adapted from Table 4·1, M. G. Kendall (1946)) are also plotted for comparison.

In order to examine the extent of validity of formula (20) and the smoothing properties of (22), it is first noted from (20) that $I_p/E(I_p)$ is approximately a χ^2 with 2 D.F. and fluctuations in any smoothed periodogram would consequently be expected to be roughly equivalent to χ^2 distributions with 4, 8, etc., degrees of freedom according to the smoothing adopted. This could be tested by means of an appropriate adaptation of the homogeneity of variances test (Bartlett, 1937). To cover deviations in both I_p and $E(I_p)$ the values of $J_p \equiv I_p/[2\pi\sigma^2 f_+(\omega)]$ were evaluated over the range for which the smoothed periodograms for the subseries had been computed. Since the subseries lengths were all 60, the case $n = 30$ should correspond to χ^2's with 4 D.F., $n = 15$ to 8 D.F.

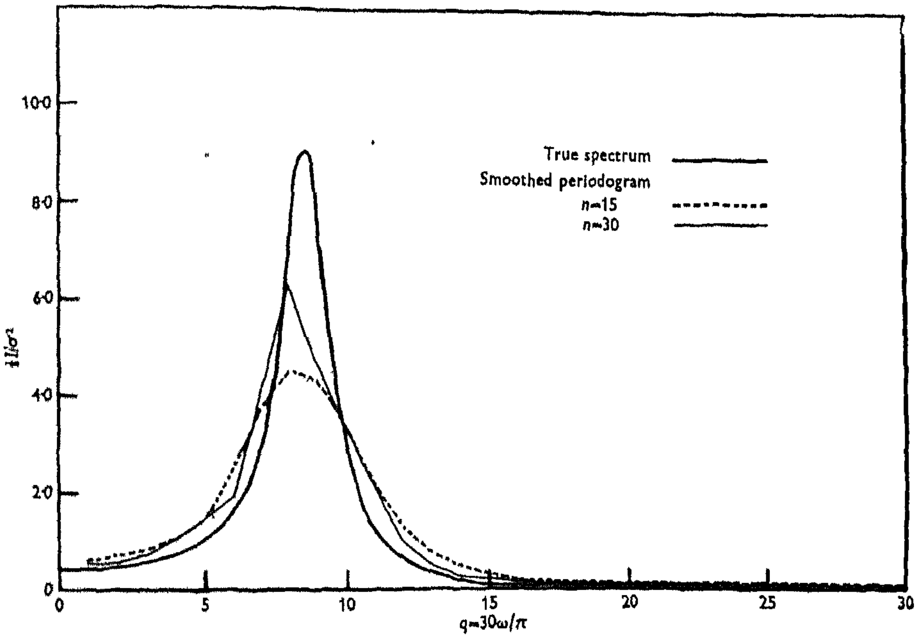

Fig. 2. Smoothed periodograms ($n = 15$ and 30) for Series III, compared with true spectrum.

In the case of Series I, this was almost completely verified. The over-all fluctuations were split up into fluctuations at any frequency ω and systematic deviations of the average at any frequency, i.e.

$$\log J_p = \log (J_p/\bar{J}_p) + \log \bar{J}_p.$$

The values of $\bar{J}_p$ could further be compared with the average over p, $\bar{\bar{J}}_p$, say, the average $\bar{\bar{J}}_p$ being tested separately. The analyses given in Tables 1 and 2 were obtained.

It will be seen that the only significant ($P = 0{\cdot}05$) departure from the asymptotic distribution was in the systematic deviations $\bar{J}_p$ for $n = 15$. This may be due to a real bias (cf. the results below for Series III), but might also be partly due to the periodogram being evaluated for integral q, which for $n = 15$ corresponds to quarter-integers for p; this would imply considerable correlation between neighbouring points even for a completely random series, whereas the above test assumes independence.

In the case of Series III it will be noticed that the amplitude of the peak of the smoothed periodogram fails to reach the 'true value', especially for $n = 15$. That this is a real

Tables 1 and 2. *Analysis of fluctuations by approximate* χ^2 *test*

Series I. $n=30$.

	D.F.	Crude χ^2	Corrected χ^2
$J_p/\bar{J}_p$	105	128·38	—
$\bar{J}_p/\bar{\bar{J}}_p$	14	17·84	—
$\bar{\bar{J}}_p$	1	0·16	—
Total	120	146·38	135·12

Series I. $n=15$.

	D.F.	Crude χ^2	Corrected χ^2
$J_p/\bar{J}_p$	105	95·62	—
$\bar{J}_p/\bar{\bar{J}}_p$	14	25·91	24·81†
$\bar{\bar{J}}_p$	1	1·21	—
Total	120	122·74	—

† Significant at $P=0{\cdot}05$ level.

Note. For a test of $\bar{\bar{J}}_p$ the *unconditional* likelihood function is required, the corresponding crude for an average of $8 \times 15 = 120$ items being

$$-240m \log \bar{\bar{J}}_p + 240m(\bar{\bar{J}}_p - 1).$$

Similarly, the correcting factor for the *total* χ^2 is $1+\frac{1}{6m}=\frac{13}{12}$ for $n=30$, $m=2$.

Tables 3 and 4

-Series III. $n=30$.

	D.F.	Crude χ^2
$J_p/\bar{J}_p$	45	27·71
$\bar{J}_p/\bar{\bar{J}}_p$	14	178·08†
$\bar{\bar{J}}_p$	1	6·01†
Total	60	211·80†

Series III. $n=15$.

	D.F.	Crude χ^2
$J_p/\bar{J}_p$	45	27·78
$\bar{J}_p/\bar{\bar{J}}_p$	14	319·83†
$\bar{\bar{J}}_p$	1	31·58†
Total	60	379·19†

† Significant at $P=0{\cdot}01$ level.

systematic error due to the short length of the equivalent length of series may be checked from the corresponding analysis of the fluctuations for the subseries (Tables 3 and 4).

This damping of the peak must, however, be expected for comparatively small n in the case of series with little-damped oscillations. It will be remembered that in the case of classical harmonic analysis with a discrete spectrum the true amplitude (with this standardization) would be infinite, whereas that of the periodogram would still be finite, and the smaller and less well resolved the fewer the observations.

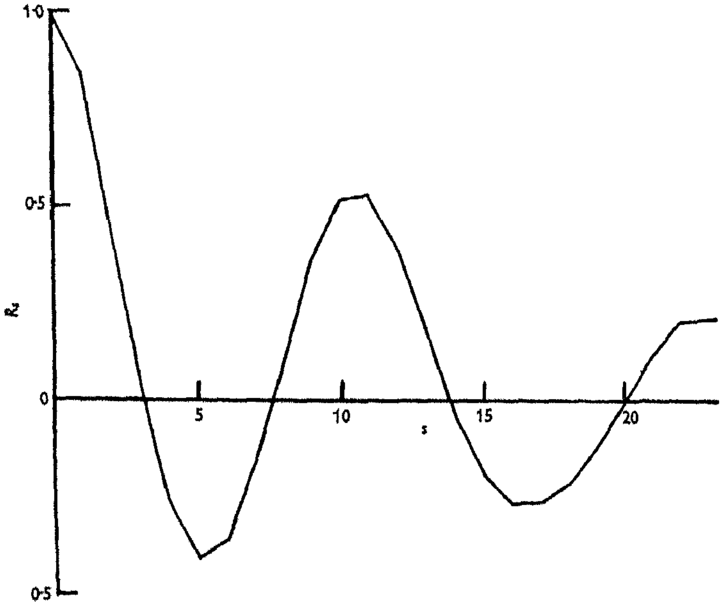

Fig. 3. Correlogram of Wolfer's sunspot numbers (Yule's graduated series, 1749–1924).

It may be concluded that with the present method of estimating a continuous spectrum by smoothing the periodogram excessive curtailing of the correlogram is inadvisable if the amplitude of the peak is to be estimated as well as its location. The estimated spectrum is in general a consistent rather than a strictly unbiased estimate, approaching the true spectrum if the series is large enough for adequate smoothing and yet n is also large enough. The value of n to take may (cf. Bartlett, 1948) be judged from the amount of damping in the correlogram.

Among subsidiary effects in the above χ^2 analysis for Series III will be noted (i) a tendency for fluctuations about $\bar{J}_p$ to be less than expected on the asymptotic theory, and (ii) a mean bias $\bar{J}_p$, especially for $n = 15$.

6. SMOOTHED PERIODOGRAM FOR WOLFER'S SUNSPOT NUMBERS

This method of smoothing the periodogram was also tried out on Wolfer's sunspot numbers as an example of an actual series, in contrast with the artificial series discussed above. This series was the first to be analysed by the autoregressive method (Yule, 1927), and for comparative purposes the present analysis uses the same (graduated) series studied by Yule, i.e. from 1749 to 1924. The observed values of R_s, for $s = 1 \ldots 23$, are given† in Appendix B

---

† There are some trivial differences between these values and those given by Yule for $s = 1 \ldots 5$, possibly due to the formulae or method of computation used.

(Table 11) and in Fig. 3. Yule's graduated values were used because previous analysis by Yule and by the author (Bartlett, 1946) respectively suggested that these values might correspond to an autoregressive model of type (13), or to the corresponding analogue defined for continuous time. However, the calculation of the more complete correlogram established the untenability of either assumption, at least in this simple form (e.g. by Quenouille's χ^2 test in the case of the discrete time model, and, in the case of continuous time, by an adaptation of it to be reported elsewhere in collaboration with P. H. Diananda).

But in spite of the absence of any simple adequate autoregressive model, the series is one, as Yule pointed out, for which classical harmonic analysis is of doubtful value, although periodogram analysis of the series has from time to time been made (Schuster, 1906; Larmor & Yamaga, 1917). In order to investigate its spectrum without a knowledge of the underlying mechanism, the smoothed periodogram (with $n = 24$) was calculated for $q = 88\omega/\pi$

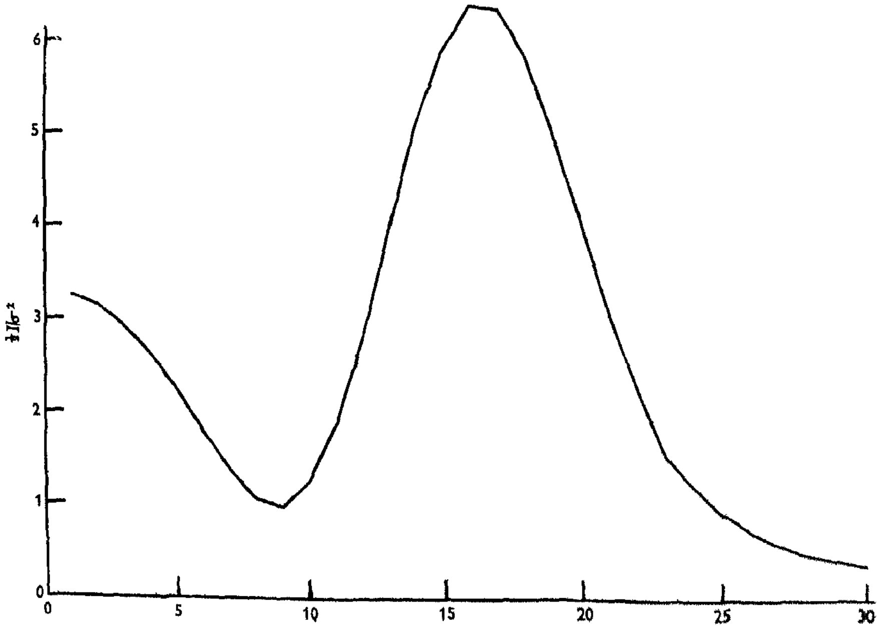

Fig. 4. Smoothed periodogram ($n = 24$) of Wolfer's sunspot numbers.

from 1 to 30 and is given in Appendix B (Table 12) and Fig. 4. There is a rise towards $\omega = 0$; in spite of possible small contributions due to the incomplete elimination of the mean or to the effect of the graduation this suggests a real component of non-oscillatory or 'Markoff' type, and indicates how a somewhat better fit might be obtained for the correlogram by addition of such a component to any theoretical model for the process. As such an empirical extension of the autoregressive model does not greatly contribute to the interpretation of the observed data it is not pursued in any detail, but it may be of interest to note Sir Gilbert Walker's somewhat similar calculations on the correlogram of a meteorological series in his search for empirical fits (Walker, 1931).

Apart from its behaviour at the origin, the spectrum over the interesting range consists of a single smooth humpbacked curve with a peak at about 11·0 years' 'period', in agreement with the usually accepted value of the period.

7. CONCLUSIONS

The discussion in this paper has emphasized that the more comprehensive theory of periodicity in time series now available throws up problems in sampling theory of considerable complexity and, up to recently, largely ignored. While other branches of mathematical statistics have developed an extensive distributional theory of sampling variability, in contrast with the earlier 'large-sample' theory, the sampling theory of time series is on the whole still in the more primitive 'large-sample' state. The estimation and assessment of accuracy of autoregressive constants and the goodness of fit test of the correlogram fall into this category.

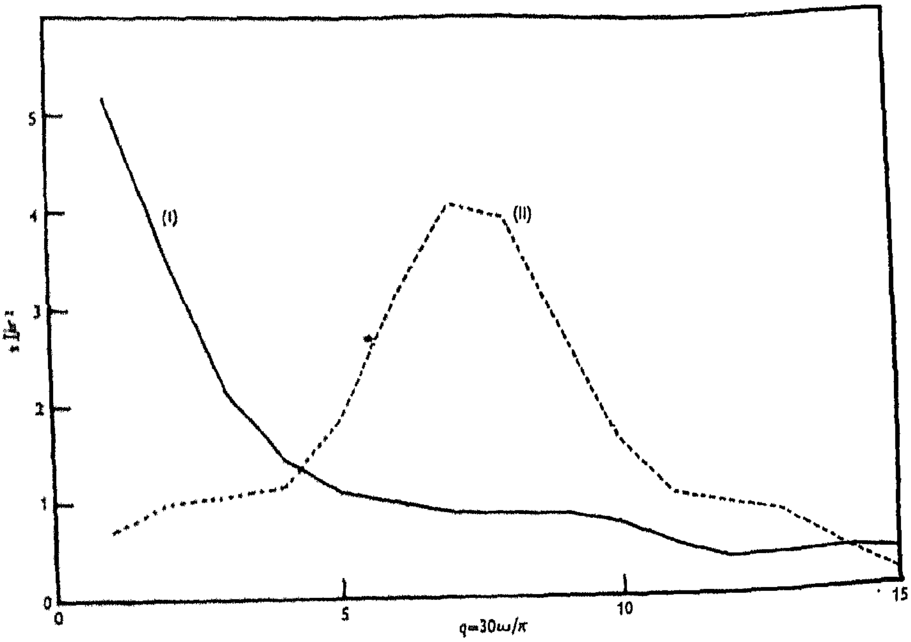

Fig. 5. Smoothed periodograms ($n = 15$) of subseries (i) and (ii) (Series I).

It has been pointed out that some method of smoothing the periodogram of a series with continuous spectrum is essential, as the classical method gives no convergence to the true spectrum. While the method recommended above also falls into the 'large-sample' category, and no precise tests of the significance or accuracy of any observed peaks have been derived, the value of the method would appear to have been established by the above examples. On short series the amount of smoothing possible will not be sufficient to prevent extensive sampling fluctuations from one 'sample' to another, as is illustrated in Fig. 5, which shows two contrasted curves for subseries (i) and (ii) (Series I), of total length 60, differing greatly in their evidence for 'periodicity'. The order of magnitude of the fluctuations is, however, known, ranging from the asymptotic formula (20) in the case of the unsmoothed periodogram to the $1/\sqrt{m}$ reduction factor when smoothing is adopted. The order of correlation and resolution for different frequencies is also known, both depending on n. This allows a rough assessment of significance of variation in amplitude to be made, and would if used earlier at least have prevented the many doubtful claims of *multiple* periods to be found in the literature (for example, by Brownlee (1917) in the case of the periodicity of measles epidemics).

SUMMARY

The problem of investigating periodicity in time series is considered in the light of theoretica developments in the harmonic analysis of stationary stochastic processes; it is for series with continuous spectra that classical periodogram analysis is often useless and misleading.

A proposed method of smoothing the periodogram, making use of the correlogram, is tested arithmetically on two of M. G. Kendall's artificial series, and a statistical analysis made of the fluctuations in the smoothed periodograms. The method is also tested on Wolfer's sunspot numbers. These empirical tests supplement a theoretical discussion of the asymptotic properties of the method.

I am indebted to Miss R. M. Dyson for most of the computational work involved in this investigation.

APPENDIX A

Smoothed periodograms for M. G. Kendall's artificial Series I and III (standardized to $\pi f_1(\omega)$ and tabulated for integral values of $q = 30\omega/\pi$)

Table 5

Series I. Total series.

q	From correlations			From covariances			True values
	$n=45$	$n=30$	$n=15$	$n=45$	$n=30$	$n=15$	
1	3·208	2·936	2·913	3·192	2·938	2·915	2·219
2	3·082	3·247	3·083	2·994	3·254	3·089	2·380
3	3·809	3·505	3·058	3·884	3·517	3·065	2·656
4	2·821	2·899	2·821	2·893	2·905	2·827	3·030
5	2·293	2·297	2·593	2·204	2·294	2·594	3·395
6	2·359	2·474	2·485	2·331	2·477	2·489	3·493
7	3·026	2·785	2·366	3·135	2·792	2·370	3·091
8	2·408	2·368	2·045	2·890	2·371	2·045	2·353
9	1·198	1·251	1·547	1·168	1·247	1·499	1·641
10	0·763	0·889	1·092	0·811	0·887	1·091	1·118
11	0·930	0·847	0·879	1·006	0·848	0·822	0·773
12	0·630	0·686	0·684	0·548	0·687	0·684	0·549
13	0·478	0·488	0·557	0·432	0·485	0·556	0·403
14	0·344	0·348	0·411	0·435	0·343	0·408	0·305
15	0·180	0·216	0·289	0·180	0·216	0·288	0·237
16	0·183	0·195	0·225	0·087	0·199	0·227	0·190
17	0·154	0·158	0·199	0·196	0·159	0·199	0·155
18	0·099	0·123	0·176	0·179	0·123	0·177	0·130
19	0·140	0·135	0·217	0·056	0·135	0·160	0·111
20	0·139	0·143	0·153	0·088	0·143	0·153	0·096
21	0·109	0·122	0·147	0·203	0·121	0·146	0·085
22	0·114	0·111	0·130	0·124	0·111	0·129	0·076
23	0·066	0·081	0·108	−0·036	0·080	0·107	0·069
24	0·066	0·067	0·093	0·097	0·070	0·092	0·064
25	0·059	0·066	0·089	0·142	0·063	0·085	0·059
26	0·052	0·062	0·085	−0·018	0·058	0·081	0·056
27	0·056	0·058	0·079	−0·008	0·054	0·076	0·054
28	0·049	0·060	0·076	0·142	0·057	0·074	0·053
29	0·055	0·057	0·077	0·065	0·055	0·075	0·052
30	0·048	0·058	0·077	−0·051	0·058	0·077	0·051

Table 6

Series III. Total series.

q	$n=30$	$n=15$	True values
1	0·523	0·630	0·473
2	0·716	0·774	0·515
3	0·937	0·882	0·596
4	1·035	1·046	0·745
5	1·290	1·555	1·023
6	1·881	2·560	1·603
7	4·113	3·758	3·031
8	6·262	4·496	6·835
9	4·672	4·309	8·079
10	3·257	3·330	3·222
11	2·171	2·138	1·337
12	0·993	1·234	0·680
13	0·534	0·721	0·399
14	0·257	0·450	0·259
15	0·223	0·291	0·180
16	0·158	0·215	0·133
17	0·156	0·162	0·102
18	0·107	0·181	0·081
19	0·107	0·152	0·067
20	0·129	0·134	0·056
21	0·119	0·129	0·048
22	0·063	0·114	0·042
23	0·054	0·083	0·038
24	0·039	0·066	0·035
25	0·055	0·074	0·032
26	0·057	0·082	0·030
27	0·051	0·070	0·029
28	0·034	0·058	0·028
29	0·047	0·064	0·027
30	0·045	0·071	0·027

Table 7

Series I. $n=30$.

q	Subseries of 60 terms									Total series
	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)	Average	
1	4·792	0·678	2·682	1·346	1·760	2·922	6·480	2·566	2·903	2·938
2	2·092	0·547	3·279	2·651	2·433	4·097	11·944	2·452	3·687	3·254
3	2·274	0·952	1·782	2·673	3·671	3·467	6·302	6·611	3·466	3·517
4	1·718	2·043	1·106	1·460	2·601	3·021	1·364	7·082	2·549	2·905
5	0·875	1·057	0·589	2·200	1·015	1·780	3·412	5·833	2·095	2·294
6	0·716	1·159	2·183	4·046	1·485	3·830	3·637	3·005	2·508	2·477
7	0·728	6·812	3·060	3·261	3·954	4·413	1·749	0·774	3·094	2·792
8	0·937	6·158	1·192	1·499	3·324	3·581	0·405	0·989	2·261	2·371
9	1·143	1·099	0·411	2·928	0·684	1·873	0·400	0·382	1·115	1·247
10	0·572	0·750	0·409	2·292	0·382	0·321	0·586	0·703	0·752	0·887
11	0·339	1·687	0·414	0·781	0·396	1·017	1·389	0·763	0·848	0·848
12	0·451	1·398	0·537	0·508	1·050	0·832	0·568	0·582	0·741	0·687
13	0·349	0·502	0·457	0·346	0·761	0·631	0·547	0·180	0·472	0·485
14	0·376	0·251	0·315	0·447	0·178	0·389	0·397	0·365	0·340	0·343
15	0·379	0·064	0·192	0·172	0·117	0·292	0·486	0·213	0·239	0·216

Table 8

Series I. $n = 15$.

q	Subseries of 60 terms									Total series
	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)	Average	
1	5·176	0·699	2·784	1·792	1·978	2·731	6·483	2·785	3·054	2·915
2	3·586	0·996	2·637	2·041	2·523	3·189	6·836	3·987	3·224	3·089
3	2·172	1·068	2·169	2·240	2·624	3·246	6·098	5·314	3·116	3·065
4	1·448	1·156	1·646	2·396	2·249	2·566	4·427	5·867	2·720	2·827
5	1·095	1·829	1·514	2·604	1·999	2·903	2·985	5·097	2·503	2·594
6	0·997	3·078	1·768	2·809	2·228	3·257	2·346	3·380	2·483	2·489
7	0·863	4·051	1·878	2·817	2·540	3·487	1·974	1·732	2·418	2·370
8	0·855	3·913	1·504	2·541	2·320	3·038	1·319	0·881	2·046	2·045
9	0·869	2·797	0·874	2·097	1·554	2·060	0·652	0·737	1·455	1·499
10	0·742	1·624	0·470	1·617	0·822	1·199	0·476	0·776	0·966	1·091
11	0·520	1·066	0·431	1·135	0·572	0·834	0·676	0·692	0·741	0·822
12	0·389	0·969	0·508	0·687	0·658	0·782	0·752	0·537	0·660	0·684
13	0·398	0·817	0·476	0·382	0·666	0·698	0·573	0·426	0·555	0·556
14	0·419	0·461	0·365	0·280	0·459	0·490	0·460	0·358	0·412	0·408
15	0·353	0·169	0·296	0·280	0·235	0·289	0·409	0·208	0·291	0·288

Table 9

Series III. $n = 30$.

q	Subseries of 60 terms					Total series
	(i)	(ii)	(iii)	(iv)	Average	
1	0·101	0·561	0·318	0·740	0·430	0·523
2	0·705	0·483	0·416	0·169	0·443	0·716
3	1·727	1·229	0·890	0·436	1·070	0·937
4	1·211	0·980	0·772	0·936	0·975	1·035
5	1·550	0·608	1·482	1·400	1·260	1·290
6	3·056	1·143	2·395	0·530	1·781	1·881
7	2·208	4·017	7·830	5·398	4·863	4·113
8	3·212	5·039	8·274	7·413	5·985	6·262
9	4·543	1·959	4·753	3·640	3·724	4·672
10	2·434	1·198	6·008	3·212	3·213	3·257
11	1·372	1·173	2·939	3·676	2·290	2·171
12	0·593	1·013	0·919	1·415	0·985	0·993
13	0·223	0·927	0·549	1·329	0·757	0·534
14	0·189	0·444	0·393	0·258	0·321	0·257
15	0·050	0·247	0·320	0·211	0·207	0·223

Table 10

Series III. $n = 15$.

q	Subseries of 60 terms					Total series
	(i)	(ii)	(iii)	(iv)	Average	
1	0·501	0·653	0·466	0·809	0·607	0·630
2	0·834	0·830	0·670	0·751	0·771	0·774
3	1·222	0·787	0·776	0·642	0·857	0·882
4	1·549	0·752	1·060	0·773	1·033	1·046
5	1·842	1·136	2·044	1·544	1·642	1·555
6	2·227	2·046	3·772	2·968	2·754	2·560
7	2·731	2·962	5·518	4·433	3·911	3·758
8	3·129	3·189	6·347	5·131	4·449	4·496
9	3·093	2·590	5·893	4·747	4·081	4·309
10	2·504	1·698	4·538	3·662	3·101	3·330
11	1·612	1·112	2·968	2·505	2·049	2·138
12	0·822	0·928	1·686	1·615	1·263	1·234
13	0·378	0·834	0·877	0·977	0·766	0·721
14	0·239	0·612	0·488	0·518	0·464	0·450
15	0·233	0·351	0·355	0·264	0·301	0·291

APPENDIX B

Table 11. *Correlogram of Wolfer's sunspot numbers* (*Yule's graduated series*, 1749–1924)

s	R_s	s	R_s	s	R_s
1	0·8408	9	0·3646	17	−0·2683
2	0·4715	10	0·5197	18	−0·2179
3	0·0470	11	0·5273	19	−0·1256
4	−0·2647	12	0·3937	20	−0·0096
5	−0·4059	13	0·1791	21	0·1143
6	−0·3601	14	−0·0390	22	0·2035
7	−0·1636	15	−0·1947	23	0·2132
8	0·1088	16	−0·2706		

Table 12. *Smoothed periodogram* ($n = 24$) *of Wolfer's sunspot numbers* (*standardized to* $\pi f_+(\omega)$ *and tabulated for* $q = 88\omega/\pi$)

q		q		q	
1	3·259	11	1·854	21	3·047
2	3·138	12	2·792	22	2·237
3	2·927	13	3·913	23	1·629
4	2·621	14	5·017	24	1·203
5	2·231	15	5·883	25	0·921
6	1·790	16	6·339	26	0·714
7	1·366	17	6·304	27	0·589
8	1·060	18	5·809	28	0·489
9	0·986	19	4·982	29	0·426
10	1·237	20	4·004	30	0·398

REFERENCES

Bartlett, M. S. (1937). *Proc. Roy. Soc.* A, **160**, 268.

Bartlett, M. S. (1946). *J. R. Statist. Soc.* (Suppl.), 8, 27.

Bartlett, M. S. (1948). *Nature, Lond.*, **161**, 686.

Brownlee, J. (1917). *Philos. Trans.* B, **208**, 225.

Cramér, H. (1942). *Ark. Mat. Astr. Fys.* **28**, no. 12.

Cramér, H. (1947). *Ann. Math. Statist.* **18**, 165.

Fisher, R. A. (1929). *Proc. Roy. Soc.* A, **125**, 54.

Karhunen, K. (1947). *Ann. Acad. Sci. Fenn.* **37**.

Kendall, D. G. (1948). *Nature, Lond.*, **161**, 187.

Kendall, M. G. (1946). *Contributions to the Study of Oscillatory Time Series.* Cambridge University Press.

Khintchine, A. (1934). *Math. Ann.* **109**, 604.

Larmor, J. & Yamaga, N. (1917). *Proc. Roy. Soc.* A, **93**, 493.

Lévy, P. (1948). *Processus Stochastiques et Mouvement Brownien.* Paris.

Moyal, J. E. (1949). *J. R. Statist. Soc.* Series B (in the Press).

Quenouille, M. H. (1947). *J. R. Statist. Soc.* **110**, 123.

Schuster, A. (1906). *Philos. Trans.* A, **106**, 69.

Slutsky, E. (1937). *Econometrica*, **5**, 105.

Slutsky, E. (1938). *Actualités sci. ind.* no. 738, p. 33. Paris.

Walker, G. (1931). *Proc. Roy. Soc.* A, **131**, 518.

Wiener, N. (1930). *Acta math., Stockh.*, **55**, 117.

Wold, H. (1938). *A Study in the Analysis of Stationary Time-series.* Uppsala.

Yule, G. U. (1927). *Philos. Trans.* A, **226**, 267.

NOTES ON CONTINUOUS STOCHASTIC PHENOMENA

By P. A. P. MORAN, *Institute of Statistics, Oxford University*

The study of stochastic processes has naturally led to the consideration of stochastic phenomena which are distributed in space of two or more dimensions. Such investigations are, for instance, of practical interest in connexion with problems concerning the distribution of soil fertility over a field or the relations between the velocities at different points in a turbulent fluid. A review of such work with many references has recently been given by Ghosh (1949) (see also Matérn, 1947). In the present note I consider two problems arising in the two- and three-dimensional cases.

RELATIONS BETWEEN CONTINUOUS AND DISCONTINUOUS PROCESSES

Stochastic variables defined for points on a plane may be considered as defined at a discrete set of points (for example, at all points with integral co-ordinates) or for a continuous domain of points. The latter is the natural approach when considering soil fertility, but in the study of the efficiency of experimental designs it is more natural to consider the fertility as varying discontinuously from plot to plot rather than within each plot. For this reason I begin by considering the relationship between continuous and discrete models of such phenomena.

First consider stationary stochastic processes in one dimension defined by variates $x(t)$, where t is 'time' and takes either integral or a continuous range of values. Continuous processes whose variate $x(t)$ has a correlation function

$$\rho(t) = \exp[-\lambda |t|] \tag{1}$$

are known (Bartlett, 1947, p. 79) to exist and to have a spectral density given by

$$W'(\theta) = \frac{2\lambda}{\pi(\lambda^2+\theta^2)}, \tag{2}$$

so that
$$\rho(t) = \int_0^\infty \cos t\theta \, dW(\theta) = 2\lambda \int_0^\infty \frac{\cos t\theta \, d\theta}{\pi(\lambda^2+\theta^2)}.$$

From such a continuous process, a discrete process can be derived in two ways. First we might consider the values of $x(t)$ only at discrete values of t ($= 0, \pm 1, \dots$ say). Such a process would have the serial correlation

$$\rho_s = \exp[-\lambda |s|] \quad (s = 0, \pm 1, \dots),$$

and could be regarded as being generated by a simple Markoff relation of the form

$$x_s = e^{-\lambda} x_{s-1} + \eta_s,$$

where $\{\eta_s\}$ is a stationary process which is not necessarily completely random but nevertheless has all its serial correlations zero.

In practice it is perhaps more realistic to consider discrete processes derived from continuous ones in another way. Suppose we write

$$X(s) = \int_s^{s+1} x(t)\, dt, \tag{3}$$

where s takes values $0, \pm 1, \ldots$ and the integral is a 'stochastic integral'. A completely rigorous discussion of this definition would be neither short nor (in the present connexion) very illuminating, and we do not enter into such considerations here. (For a general discussion of such questions see Lévy, 1948.) It is clear that $\rho(t)$ is continuous at $t = 0$ and consequently $x(t)$ is 'stochastically continuous' (Bartlett, 1947, p. 77). To discuss the correlational and spectral properties of $X(s)$ we may therefore argue as follows. We approximate to

$$X(s) = \int_s^{s+1} x(t)\,dt$$

by a sum, and to find $\operatorname{var} X(s)$ and $\operatorname{cov}\{X(s), X(s+k)\}$, we take the expectations of the sums and proceed to the limit. In this way we find

$$\begin{aligned}\operatorname{var} X(s) = E\{X(s)\}^2 &= \int_0^1 da \int_0^1 db \exp[-\lambda\,|\,a-b\,|]\\ &= \frac{2}{\lambda^2}\{\lambda - 1 + e^{-\lambda}\}.\end{aligned}$$

Similarly, for $k \geqslant 1$

$$\begin{aligned}E\{X(s)\,X(s-k)\} = E\{X(s)\,X(s+k)\} &= \int_0^1 da \int_0^1 db \exp[-\lambda\,|\,k+b-a\,|]\\ &= \frac{1}{\lambda^2} e^{-\lambda k}(e^{\lambda}-1)(1-e^{-\lambda}).\end{aligned}$$

Thus ρ_k $(k \geqslant 1)$, the serial correlation of $X(s)$ and $X(s+k)$, is given by

$$\rho_k = \rho_{-k} = \frac{e^{-\lambda k}(e^{\lambda}-1)(1-e^{-\lambda})}{2(\lambda-1+e^{-\lambda})} = A\,e^{-\lambda k}, \quad \text{say.} \tag{4}$$

Then as $\lambda \to 0$, $\rho_k \to 1$ and as $\lambda \to \infty$, $\rho_k \to 0$ as we expect. Now, using the fact that

$$\sinh\lambda = \tfrac{1}{2}(e^{\lambda} - e^{-\lambda}) > \lambda \quad \text{for} \quad \lambda > 0,$$

it can be easily shown that

$$(e^{-\lambda}-1)(1-e^{-\lambda}) > 2(\lambda - 1 + e^{-\lambda}),$$

and therefore

$$A > 1.$$

On the other hand, it is easily verified algebraically that $\rho_1 = A\,e^{-\lambda} < 1$. Formula (4) is not the sort of correlation *function* which would be obtained for a process which is the solution of a simple stochastic difference equation of Markoff type because $A > 1$, but we may construct a simple mechanism which would generate a process having the above correlational properties. Write $r = e^{-\lambda}$. Then the serial correlation generating function of the process $X(s)$ is

$$\begin{aligned}\sum_{-\infty}^{\infty} \rho_k z^k &= 1 + A\sum_1^{\infty} r^k z^k + A\sum_1^{\infty} r^k z^{-k}\\ &= 1 + \frac{Arz}{1-rz} + \frac{Arz^{-1}}{1-rz^{-1}}\\ &= \frac{1 + r^2 - 2Ar^2 + (A-1)rz + (A-1)rz^{-1}}{(1-rz)(1-rz^{-1})}\\ &= \frac{(\alpha+\beta z)(\alpha+\beta z^{-1})}{(1-rz)(1-rz^{-1})},\end{aligned}$$

where

$$\alpha, \beta = \tfrac{1}{2}\{(1-r)^2 + 2Ar(1-r)\}^{\frac{1}{2}} \pm \tfrac{1}{2}\{(1+r)^2 - 2Ar(1+r)\}^{\frac{1}{2}}.$$

α and β can be easily verified algebraically to be real. Then if $\{\xi_n\}$ is a sequence of independent random variables with zero means and the same standard deviation, the process $\{Y_n\}$ generated by the relation

$$Y_{n+1} = rY_n + \alpha\xi_n + \beta\xi_{n-1},$$

will have the same correlational properties as $\{X_n\}$.

We now consider what happens when we generalize the above to 'processes' with more than one 'time'. As the idea of a continuing 'process' is now less applicable we shall call such a model a 'spatial stochastic system'. We consider such a system to be defined if for any set of points $P_1, \ldots, P_k$ in such a plane (or higher dimensional space) there is given a set of random variables $x_1, \ldots, x_k$ and a corresponding joint distribution function $F(x_1, \ldots, x_k)$ which satisfies the customary consistency conditions that the joint distribution of any set of x's, $x_1, \ldots, x_p$ $(p<k)$ is obtained from the joint distribution of $x_1, \ldots, x_k$ by integrating out $x_{p+1}, \ldots, x_k$. This condition corresponds to the Chapman-Kolmogoroff equation in the theory of processes with a single 'time'. If the distribution function $F(x_1, \ldots, x_k)$ is invariant under any translation of the set of points $P_1, \ldots, P_k$ we call the system 'stationary', and if, in addition, it is invariant under any rotation we call it 'isotropic'. If we only know that the first- and second-order moments of $x_1, \ldots, x_k$ are invariant under such a translation (or rotation), we say the system is stationary (or isotropic) to the second order. In what follows we consider stationary processes, but the results obviously hold under the weaker condition also.

Suppose the variates defined so that each has zero expectation and variance σ^2. Considering first systems in two dimensions we take two parameters t and u to correspond to the 'times', and we then have a correlation function

$$\rho(p,q) = \text{correlation}\{x(t,u),\, x(t+p, u+q)\}.$$

It follows that

$$\rho(p,q) = \rho(-p,-q),$$

but it is not necessarily true that

$$\rho(p,q) = \rho(p,-q) \quad \text{or} \quad \rho(-p,q).$$

The latter relations would be true if the system were isotropic, but they are not a sufficient condition for isotropy.

The natural generalization of processes with a correlational function (1) would be a system whose correlational function is

$$\rho(p,q) = \exp[-\lambda(p^2+q^2)^{\frac{1}{2}}]. \tag{5}$$

As will be seen later it is easy to show that such systems exist. Now suppose we derive from a continuous system having (5) as correlation function, a two-dimensional discrete system defined by a system of variables X_{lm}, where l, m take integral values, and X_{lm} is defined by

$$X_{l,m} = \int_l^{l+1}\int_m^{m+1} x(t,u)\,dt\,du.$$

Following our previous argument we might expect that we should have a correlation function which generalizes (5), i.e. of the form (5) multiplied by a constant. That this is not true can be seen as follows. If the result were true, the covariance of X_{lm} and $X_{l+p,m+q}$ would be of the form

$$K\exp[-\mu(p^2+q^2)^{\frac{1}{2}}],$$

where K is some constant and μ is not necessarily equal to λ. It would then follow that

$$T = \exp[\mu(p^2+q^2)^{\frac{1}{2}}]\operatorname{cov}(X_{l,m}, X_{l+p,m+q})$$

would be independent of p and q. Taking in particular the case $q = 0$, this is equal to

$$T = \int_0^1 dt_1 \int_0^1 du_1 \int_0^1 dt_2 \int_0^1 du_2 \exp\{\mu p - \lambda[(p+t_2-t_1)^2 + (u_2-u_1)^2]^{\frac{1}{2}}\}.$$

Then
$$\frac{\partial T}{\partial p} = \int_0^1 dt_1 \int_0^1 du_1 \int_0^1 dt_2 \int_0^1 du_2 \exp\{\mu p - \lambda[(p+t_2-t_1)^2 + (u_2-u_1)^2]^{\frac{1}{2}}\}$$
$$\times \left\{\mu - \frac{\lambda(p+t_2-t_1)}{\{(p+t_2-t_1)^2 + (u_2-u_1)^2\}^{\frac{1}{2}}}\right\}.$$

If $\mu \geqslant \lambda$ the second factor is greater than zero over the whole range of integration with the possible exception of the points where $u_2 = u_1$, and so the whole integral is non-zero, which contradicts the hypothesis. On the other hand, if $\mu < \lambda$ we can choose p so large that over the whole range of integration the second factor is negative, again contradicting the hypothesis. Thus this generalization of the previous argument breaks down.

It follows that in cases where it is necessary to derive a discrete system from a continuous one, in the above manner, it would be more convenient, if perhaps occasionally less realistic, to consider systems with a correlation function

$$\rho(p, q) = \exp[-\lambda|p| - \mu|q|]. \tag{6}$$

It is easy to see that in this case the discrete process $X_{l,m}$ has a correlation function

$$\rho_{kl} = A_1 A_2 \exp[-\lambda|k| - \mu|l|],$$

where
$$A_1 = \frac{(e^{\lambda}-1)(1-e^{-\lambda})}{2\{\lambda - 1 + e^{-\lambda}\}},$$

and
$$A_2 = \frac{(e^{\mu}-1)(1-e^{-\mu})}{2\{\mu - 1 + e^{-\mu}\}},$$

when neither k nor l is equal to zero. When one is equal to zero the formula is modified.

Cases where correlation functions of this kind can plausibly arise are those in which we might attempt to study the relative efficiencies of various experimental designs assuming that soil fertility is a random variable with a spatial correlation of the form (6).

The existence of spatial stochastic systems with prescribed correlation functions follows from the two- (or more) dimensional analogue of Khintchine's theorem. This asserts that a necessary and sufficient condition for the existence of a spatial stochastic system with a correlation function $\rho(p, q)$ is that there exist a function $W(x, y)$, non-decreasing in x and y, such that

$$\rho(p, q) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i(px+qy)} \, dW(x, y)$$

and
$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dW(x, y) = 1.$$

The sufficiency of this condition is easily proved by a simple generalization of Khintchine's method (1934) of constructing a stochastic process whose correlation function is the desired one, whilst the necessity follows from a theorem of H. Cramér (1939) (see also Lévy, 1948). $W(x, y)$ can then be found in terms of $\rho(p, q)$ by a Fourier inversion formula and in particular if the system is isotropic, $W_{xy}(x, y)$, if it exists, will be a function of $(x^2 + y^2)$ only.

We can thus prove that $\exp[-\lambda(p^2+q^2)^{\frac{1}{2}}]$ is a possible correlation function (as proved in Matérn, 1947, p. 27). It is clearly sufficient to take $\lambda = 1$. We then verify that $W(x, y)$ can be taken as

$$\frac{1}{2\pi} \int_{-\infty}^{x} \int_{-\infty}^{y} \frac{du\, dr}{(1 + u^2 + r^2)^{\frac{3}{2}}}.$$

Consider

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{e^{i(px+qy)}\,dx\,dy}{(1+x^2+y^2)^{\frac{3}{2}}}.$$

Write

$$x = t\cos\alpha,\quad y = t\sin\alpha,\quad p = a\cos\beta,\quad q = a\sin\beta.$$

The above is then equal to

$$\frac{1}{2\pi}\int_0^{\infty}dt\int_0^{2\pi}d\alpha\,\frac{t\,e^{ita\cos(\alpha-\beta)}}{(1+t^2)^{\frac{3}{2}}}$$

$$=\frac{1}{2\pi}\int_0^{\infty}dt\int_0^{2\pi}d\alpha\,\frac{t\,e^{ita\cos\alpha}}{(1+t^2)^{\frac{3}{2}}}.$$

Now

$$J_0(ta)=\frac{1}{2\pi}\int_0^{2\pi}e^{ita\cos\alpha}\,d\alpha$$

(Whittaker and Watson, 1935, p. 364), and so the above integral is equal to

$$\int_0^{\infty}\frac{tJ_0(ta)}{(1+t^2)^{\frac{3}{2}}}\,dt.$$

Now it is known that

$$\int_0^{\infty}\frac{tJ_0(ta)\,dt}{(1+t^2)^{\frac{3}{2}}}=\frac{a^{\frac{1}{2}}K(a)}{2^{\frac{1}{2}}\Gamma(\frac{3}{2})}.$$

This formula can be obtained by putting $\nu = 0$, $\mu = \frac{1}{2}$ in formula (7·11·6) of Titchmarsh (1937, p. 201). But

$$K_{\frac{1}{2}}(z)=\left(\frac{\pi}{2z}\right)^{\frac{1}{2}}e^{-z},$$

and so the above integral equals $e^{-a} = \exp[-(p^2+q^2)^{\frac{1}{2}}]$. The above method of argument can be easily generalized to three dimensions and $\exp[-\lambda(p^2+q^2+r^2)^{\frac{1}{2}}]$ can be shown to be a possible correlation function. This type of function gives a satisfactory representation of the correlation between velocities in certain cases of turbulence, and it may here be pointed out that the spectral theory of turbulence has recently been developed in a very elegant form by Batchelor (1949).

By using an inversion formula we also see that in two or more dimensions, if the system is isotropic, $W(x, y, \ldots)$ will be a function of $x^2+y^2+\ldots$ only, and by integrating first over all directions, we see that for isotropic processes the correlation function will always be representable as a Fourier-Bessel-Stieltjes integral of the type $\int_0^{\infty}J_n(ut)\,t\,dF(t)$.

TEST OF THE EXISTENCE OF TWO DIMENSIONAL STOCHASTIC SCHEMES

Another problem arises in practice when we are given a set of variates $X_{i,j}$ (i, j taking integral values), and we wish to decide whether there is any evidence that these variates are spatially correlated. Such a case can arise, for example, in uniformity trials in agricultural research. This is the two-dimensional analogue of the problem of testing the significance of serial correlation coefficients on which a great deal has been written (for references see Moran, 1948*a*). We give here a simple test for correlation between nearest neighbours which generalizes a method of a previous paper (Moran, 1948*a*).

We suppose that we have mn independent variates x_{ij} ($i = 1, \ldots, m$; $j = 1, \ldots, n$), and we define what seems to be a natural definition of a correlation coefficient between x's which are nearest neighbours. Write

$$mn\bar{x}=\sum_{i,j}x_{ij}\quad\text{and}\quad z_{ij}=x_{ij}-\bar{x},$$

and
$$\begin{aligned} r_{11} &= \left(\frac{mn}{2mn-m-n}\right)\frac{\sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n-1} z_{ij}z_{i,j+1} + \sum\limits_{i=1}^{m-1}\sum\limits_{j=1}^{n} z_{ij}z_{i+1,j}}{\sum\limits_{i,j} z_{ij}^2} \\ &= \left(\frac{mn}{2mn-m-n}\right) I. \end{aligned}$$

The initial factor is conventionally introduced because there are mn terms in the denominator and $2mn-m-n$ in the numerator. In large samples r_{11} could therefore be considered an appropriate estimator of a presumed correlation coefficient between nearest neighbours. As we are here only concerned with a test for randomness it is sufficient to consider a test using I alone. If the x_{ij} are independently distributed in the same distribution, the z_{ij} are all on an equal footing, and if their distribution had a finite second moment, the correlation between any two z's would be $(mn-1)^{-1}$. Without assuming, however, that any moments exist, we have

$$\begin{aligned} E(I) &= E\left\{\frac{\sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n-1} z_{ij}z_{i,j+1} + \sum\limits_{i=1}^{m-1}\sum\limits_{j=1}^{n} z_{ij}z_{i+1,j}}{\sum z_{ij}^2}\right\} \\ &= (2mn-m-n)\,E\left\{\frac{z_{11}z_{12}}{\sum z_{ij}^2}\right\} \\ &= \left(\frac{2mn-m-n}{mn(mn-1)}\right)E\left\{\frac{(\sum z_{ij})^2-\sum z_{ij}^2}{\sum z_{ij}^2}\right\} \\ &= -\frac{2mn-m-n}{mn(mn-1)}, \end{aligned}$$

because $\sum z_{ij} = 0$. (The $\sum$ symbol without suffixes is used above to indicate summation over all values of i and j.)

To evaluate the second moment of I we have to assume further that all the x_{ij} are distributed normally. There is then no real restriction in taking the variance of x_{ij} to be unity. The denominator of I is then distributed as χ^2 with $mn-1$ degrees of freedom, and since I is the ratio of a quadratic form in the z_{ij} to a quantity $\sum z_{ij}^2$ distributed as χ^2 it follows that I is itself distributed independently of $\sum z_{ij}^2$. We then have for any positive integer p

$$E(I^p) = \frac{E(\text{numerator of } I)^p}{E(\sum z_{ij}^2)^p}.$$

But from the properties of the χ^2 distribution we have

$$E(\sum z_{ij}^2)^2 = (mn-1)(mn+1).$$

We now have to find the expectation of

$$\left(\sum_{i=1}^{m}\sum_{j=1}^{n-1} z_{ij}z_{i,j+1} + \sum_{i=1}^{m-1}\sum_{j=1}^{n} z_{ij}z_{i+1,j}\right)^2.$$

On multiplying this out we find terms of three different kinds.

(1) Terms of the form $z_{11}^2 z_{12}^2$. The number of such terms is clearly equal to the number of joins of nearest neighbours on the lattice, i.e. to $(2mn-m-n)$. Moreover, by considering the characteristic function of the joint distribution of the z's we see that

$$E(z_{11}^2 z_{12}^2) = (1+2\rho^2)\,\sigma^4,$$

where ρ is the correlation coefficient $(mn-1)^{-1}$ between any two z's and

$$\sigma^2 = \operatorname{var}(z) = (mn-1)/mn.$$

The contribution of terms of this kind to the expectation of the numerator is therefore

$$(2mn-m-n)\,(1+2(mn-1)^{-2})\,\sigma^4$$

$$= \frac{(2mn-m-n)\,(m^2n^2-2mn+3)}{(mn-1)^2}\,\sigma^4.$$

(2) The second type of term is of the form $z_{11}z_{12}^2z_{13}$, and it is easy to see (Moran, 1947, p. 323) that the number of such terms occurring in the square of the numerator is

$$2\{m(n-2)+n(m-2)+4(m-1)\,(n-1)\} = 4\{3mn-3m-3n+2\}.$$

Moreover, the expectation of such a term is $(\rho+2\rho^2)\,\sigma^4$, and so the total contribution is

$$-\frac{4(3mn-3m-3n+2)\,(mn-3)}{(mn-1)^2}\,\sigma^4.$$

(3) The third type of term is of the form typified by $z_{11}z_{12}z_{33}z_{34}$ and corresponds to two joins on the lattice without common points. The number of such terms in the expansion of the numerator is (Moran, 1947, p. 323)

$$4m^2n^2-4m^2n-4mn^2+m^2+n^2-12mn+13m+13n-8,$$

and the expectation is $3\rho^2\sigma^4$, so that the total contribution is

$$\frac{3}{(mn-1)^2}\{4m^2n^2-4m^2n-4mn^2+m^2+n^2-12mn+13m+13n-8\}\,\sigma^4.$$

Adding the above contributions and dividing by $(mn-1)\,(mn+1)$, we find

$$E(I^2) = \frac{2m^3n^3-m^3n^2-m^2n^3-4m^2n^2+2m^2n+2mn^2-2mn+3m^2+3n^2}{m^2n^2(mn-1)\,(mn+1)},$$

and var I is best found from the formula

$$\operatorname{var}(I) = E(I^2)-[E(I)]^2.$$

Higher moments of the distribution of I could be calculated in the same way using the frequencies of various combinations of joins given in Moran (1948*b*), but this would be very arduous. If, however, the mean of the distribution of the x's is known exactly, and r_{11} and I defined using deviations from this mean, the formulae are considerably simpler. In both cases it is easy to show that the distribution of I tends to normality as m and n increase. It should also be noticed that a test, based on rearrangements, for randomness in an array of this kind has been given by M. N. Ghosh (1948).

REFERENCES

BARTLETT, M. S. (1947). *North Carolina Lectures on Stochastic Processes.*
BATCHELOR, G. K. (1949). *Proc. Roy. Soc.* A, **195**, 513–32.
CRAMÉR, H. (1939). *Trans. Amer. Math. Soc.* **46**, 191–201.
GHOSH, B. (1949). *Bull. Calcutta Stat. Ass.* **2**, 11–28.
GHOSH, M. N. (1948). *Sankhyā*, **9**, 39–46.
KHINTCHINE, A. (1934). *Math. Ann.* **109**, 604–15.
LÉVY, P. (1948). *Processus Stochastiques.* Paris.
MATÉRN, B. (1947). *Metoder att uppskatta noggrannheten vid linje-och provytetaxering.* Stockholm.
MORAN, P. A. P. (1947). *Proc. Camb. Phil. Soc.* **43**, 321–8.
MORAN, P. A. P. (1948*a*). *Biometrika*, **35**, 255–60.
MORAN, P. A. P. (1948*b*). *J.R. Statist. Soc.* B, **10**, 243–51.
TITCHMARSH, E. C. (1937). *Theory of Fourier Integrals.* Oxford.
WHITTAKER, E. T. & WATSON, G. N. (1935). *Modern Analysis.* Cambridge.

ACCIDENT PRONENESS

By MAJOR GREENWOOD, F.R.S.

[*Editorial note.* This paper was probably Major Greenwood's last completed work. It was submitted for publication on the day of his death, 5 October 1949. It is hoped to include some account of his life and work in the next issue of *Biometrika*.]

An eminent psychologist has written: 'When during the last war two statisticians had their curiosity aroused by some accident figures in munition works, history was made' (Smith, 1943, p. 183). Dr May Smith did not mean that these two statisticians were the first to remark that some people had clumsier fingers and (or) less alert brains than others, a fact no doubt commented on by the overlookers in manufactories of flint arrows in the old stone age; she meant that they were the first (so far as at present known) to propose an arithmetically simple test, based on the accident frequency-distribution which would result if those exposed to the risk of accident, in an environment which did not change, had some human characteristic, later to be christened Accident Proneness, which was a continuous variable adequately represented by a Pearsonian curve of type III. Were this true, then the distribution of 0, 1, 2, etc., accidents would be a negative binomial of the form

$$\left(\frac{c}{c+1}\right)^r\left[1+\frac{r}{c+1}+\frac{r(r+1)}{2!(c+1)^2}+\frac{r(r+1)(r+2)}{3!(c+1)^2}+\dots\right], \tag{1}$$

where r and c can be deduced from the first and second moments of the distribution (Greenwood & Yule, 1920).

This simple test interested psychologists and suggested to them the possibility of detecting persons prone to accidents *before* they had had any accidents. The first thing to do was to apply suitable tests to persons who had been or were about to be exposed to the risk of accidents, and to learn whether the persons with high (or low) marks in the tests had low (or high) accident rates.

The first results of such an inquiry appeared in 1926 (Farmer & Chambers 1926), and several other reports have since been published by the same and other investigators; indeed, there is a large 'literature' of the subject. I do not propose to epitomize the literature, but believe that the following paragraphs fairly estimate what is now established.

A large number of tests have been tried. Some of these are primarily measures of corporeal dexterity, others—at the other end of the scale—explore, mainly, the conative or the cognitive sides of human nature or, if the reader prefers, of the psyche. Results show conclusively enough that there is a statistically significant correlation (measured, of course, in various ways) between performance in some of the tests and accident scores. We may therefore hold not only that 'accident proneness' is a human characteristic, but that it is measurable and that it is reasonable to believe that if persons whose scores in tests found to correlate with accident scores in these researches were excluded from occupations involving special risk of accidents, the accident rate in such occupations would be reduced. But so far no test, or battery of tests, has been found which would exclude the highly accident-prone *without* excluding a considerable proportion of entrants who are not specially 'bad risks'.

This difficulty is not peculiar to the study of accident proneness. It is found in all attempts at *a priori* selection of persons suitable for occupations. But it is a *serious* difficulty because, until it is overcome, it means that we cannot by purely arithmetical methods discover the true frequency distribution of the variable we call accident proneness, because we only have the distribution of accidents. We can only guess. We know empirically that a very large proportion of accident distributions give *some* of the criteria of a negative binomial, viz. the appropriate first and second moments of such a distribution. We know nothing more. E. M. Newbold (1927, pp. 504–5) in her classical memoir found in eleven instances that the proneness distribution was either in the type VI_J or type I_J area for ten cases while for one case it was in the 'impossible area', when the first four moments of the accident distributions were used. She justly remarks that, owing to the large probable errors of the higher moments and the fact that type III is at the dividing line of two areas, all one could say was that type III was a reasonable choice. Readers of *Biometrika* do not need to be told that because an aggregation of frequency distributions, made by summing the zeroes, the 1's, the 2's, etc., of the separate distributions, give a negative binomial, the finding is no proof that proneness was responsible. We should necessarily reach a negative binomial if the items so aggregated were Poissons—unless, of course, they were identical Poissons (Pearson, 1917, p. 139). Some beginners have overlooked this. It may, perhaps, be worth noting that if the components aggregated—all 0's together, all 1's together, etc.—were Poissons, the variance of the resulting negative binomial will be smaller than it would have been if the several components were themselves negative binomials with the same means.

There is an oddity about equation (1) as practically applied which has not been noticed. As the constants of the proneness distribution are deduced from those of the accident distribution, they must vary from factory to factory if the external risk varies. But, by hypothesis, proneness is a character of the individual, a psycho-physiological property which exists whether he is subjected to the hazard of accidents or not. This did not affect the use of the method made by Greenwood & Yule, who assumed that in each factory the risk was constant, but it does make comparison of factory with factory difficult. Suppose then that we put the fundamental Poisson in the form

$$e^{-kx}(1+kx/1!+k^2x^2/2!+k^3x^3/3!\ldots), \tag{2}$$

where k is constant for any one factory or department of a factory, but different for different factories or departments. Then our accident distribution will become

$$c^r/(c+k)^r\,(1+kr/(c+k)\,1!+k^2r(r+1)/(c+k)^2\,2!\ldots). \tag{3}$$

At first sight, this is merely trivial; if we write in (2) $y = kx$, we simply have a straight Poisson in y, and if we write ck for c in (3) we are back to (1). But it does make *some* comparison between different sets of data possible and admits of a process akin to the standardization of death-rates. Suppose we take an accident distribution, *any* accident distribution provided it is effectively a negative binomial distribution with a mean m, and adopt this as the standard form, viz. assume its k to be unity. We take now another accident distribution having a different mean m' which again is a negative binomial. Now put $k = m'/m$ and substitute for the c of the standard form $c/k = C$. We shall have again a negative binomial having the required mean. Does this effectively graduate the experience? Trials on some of Newbold's data (Newbold, 1926) showed that this very naïve plan did not give bad results, but they were not at all convincing because the range of means was narrow. She had, however, one pair of observations from the same factory, a manufactory of sweets, where the length of observations was the same for both sets but the means of accidents quite different. The sugar boilers had an accident mean of 2·50 and the other constants were $c = 0·60902$, $r = 1·52255$. The department of chocolate moulding had an accident mean of 3·94, with $c = 0·36462$ and $r = 1·4366$. Taking the sugar boilers as standard, k for moulders is 1·576, and the c should be 0·3864, which is not widely different from its empirical value 0·36462, while 1·4366 is within sight of 1·52255. Of course this is only what Darwin might have called a Tom Fool Experiment, but there may be something in the *idea*. At least, if the results of 'standardization' are widely discrepant from the observed results, it suggests heterogeneity.

An obvious cause of heterogeneity is selection. The second column of Table 1 gives the observed values of accident frequencies in 166 London Transport bus drivers (first year of observation) studied by Farmer & Chambers (1939). The second column gives the expected numbers when c and r were deduced from the observed mean, 1·81, and variance, 2·91092. The c computed by 'standardization' on the sugar boilers was half the observed value, 0·8397 instead of 1·6519, and the observed r twice the 'invariant' 1·5225. The observed accident variance was 2·91092, but 'standardization' required it to be 3·9727. The 'expected' frequencies, although not so good as those directly calculated, were not bad (third column), but evidently the variance is hopelessly large. The results *ought* to be bad. Bus drivers are a highly select population (I shall refer to it when I return from this digression); the variability in proneness of the exposed to risk must be less than that of the confectionery workers, who were certainly not so stringently selected.

Table 1. *Accidents of* 166 *bus drivers in first year of observation* (Farmer & Chambers)

No. of accidents	Observed	Expected 1	Expected 2	Expected 3
0	45	40·2	50·3	45·4
1	36	45·4	41·6	41·7
2	40	34·2	28·6	31·2
3	19	21·5	18·2	20·7
4	12	12·1	11·2	12·7
5	8	6·4	6·7	7·1
6 or more	6	5·5	9·4	7·2

Expected 1 is obtained by fitting from the observed mean and variance of the data. $P = 0·38$ for 4 degrees of freedom. Expected 2 is fitted by 'standardization'; agreement in total and mean are forced. $P = 0·19$ for 5 degrees of freedom. Expected 3 is fitted from the curtailed gamma function (the gamma function which is *assumed* to represent the proneness of the standard population) and gives the accident mean of the observed data. $P = 0·64$ for 5 degrees of freedom.

It is quite easy to determine the effects of selection, *if* proneness is obliging enough to have a gamma function distribution, thanks to the existence of Pearson's *Tables of the Incomplete Gamma Function*.* Taking the sugar boilers as standard population, the mean of the bus drivers is reproduced by a curtailment of about 16% of the frequency. Using this, the process just indicated leads to the third set of 'expected' accidents which is an excellent fit. The arithmetic amused me and may, perhaps, amuse the reader, but I do not suggest that it is of much practical value; I have in fact tacitly dropped the notion of a k constant for any one factory department, but changing from department to department. But if we take as standard population any population with a greater mean and variance than the population to be standardized, then—provided the accident distribution of the latter is fairly well fitted by a negative binomial—we should expect the process to give us a fairly good fit. Common sense, however, tells us both selection of workers and differences of environmental risk are involved. After all, we *know* nothing more than the accident distributions given by samples, and not very large samples, of two populations; too much room for guessing remains even for those whose algebra is better than mine.

---

* The terms in (1) were obtained by integration of successive complete Γ functions; these must be replaced by the appropriate $I(u, p)$'s.

Since the publication of Newbold's memoir, field workers have not bothered about a determinist approach on these lines and have used Newbold's stochastic results. To these, with certain exceptions to be noted (see p. 28), little has been added; it has, however, been pointed out that the algebra would have been simplified by working with factorial instead of power moments.

Newbold defined *statistical* proneness as meaning that in unit population $\overset{\infty}{\underset{0}{S}} f_{\lambda_r} = 1$, the accidents happening to the subfrequency f_{λ_r} will be distributed by a Poisson law with parameter λ_r, so that the 'accident universe' will be an aggregate of Poissons. What we have is a sample of persons from a universe so defined; from such a sample we can stochastically estimate the parameters of the 'universe'. By definition, the best available estimate of the mean $\bar{\lambda}$ is the mean of accidents, $\bar{A}$, and it is easy to see that if N (the number exposed to risk) is so large that $N-1$ does not differ appreciably from N, the best estimate of the variance of λ is the variance of accidents less the mean. In the particular case leading to equation (1) this is obvious. The variance of the gamma function is r/c^2, that of the accident distribution, $r/c^2+r/c$. If $A_i = \lambda_i + \epsilon_i$, where ϵ_i is a sampling error and if λ_i and ϵ_i are not correlated, we have

$$r_{A\lambda} = \frac{\sigma_\lambda}{\sigma_A} = \sqrt{\left(1 - \frac{\bar{A}}{\sigma_A^2}\right)}.$$

If some other variable, x, say the scoring of a person in a test, is correlated with his accident score, then

$$r_{\lambda x} = \frac{r_{xA}}{\sqrt{\left(1 - \frac{\bar{A}}{\sigma_A^2}\right)}}.$$

These are the relations which have been most used by field workers, whose object has been to reach a test, or battery of tests, which correlates highly with λ.

The difficulty of non-comparability of data collected under different environments of course remains. For instance, suppose correlations of test scores and accidents vary significantly between different studies. Is that due to selection? Dr Irwin has pointed out to me how that question could be answered, if one had rather more adequate numerical data. But the variation of what I have called external risk remains. *Any* statistics of accidents must contain a number of events which have no connexion with the personal qualities of the exposed to risk. We may safely infer from what we already know that the proportion is not large, but it must vary with the occupation. I suppose a minute—and objective—record of *every* accident entered on the statistical statement *might* enable us to eliminate these, but it would be a statistician's paradise in which such information was available.

Between 1927 and 1941 no important additions were made to the field-worker's statistical tools, but at least two mathematical statisticians pointed out that a negative binomial distribution of accidents, although a necessary, was not a sufficient condition for the validity of the proneness hypothesis (Irwin, 1941, pp. 102–5). For reasons given by Irwin (pp. 106–7; see also Greenwood, 1941, pp. 107–8), the alternative hypotheses probably need not be entertained when the question is of the kind of industrial accidents which have furnished the data used, so far, by all field workers.

In 1941, Yule showed that if proneness, λ, is not increased by increased length of exposure to risk, then if we write $\bar{A}_k$ for the mean accident rate over a time of exposure k, then ${}_k v_\lambda^2 = \frac{{}_k\sigma_A^2 - \bar{A}_k}{\bar{A}_k^2}$ is invariant with respect to time of exposure; then ${}_k r_{A\lambda}$ is given by

$$ {}_k v_\lambda = {}_k r_{A\lambda} \frac{{}_k\sigma_A}{\bar{A}_k}. $$

Chambers, however, found from the data of bus drivers that ${}_\lambda v_\lambda$ decreased with length of exposure, i.e. is changed by experience. For bus drivers in the first year ${}_k v_\lambda$ was 0·578, in the fifth year 0·444.

On the practical side much has been done in the last 20 years, especially by Farmer & Chambers. In their first report (1926) the subjects were girls employed in covering or packing sweets, dockyard and R.A.F. apprentices; in their last report (1939) bus drivers in the London area. These reports have certainly verified the existence of accident proneness and proved that this quality is correlated with the scoring of certain sensori-motor tests. One may conclude, as Irwin put it (1941, p. 107): '(1) That, apart from differences between drivers, accidents are certainly occurring at random. (2) The absence of a significant difference from year to year rules out the hypothesis of increasing liability to accident due to previous accidents.' The reader will remark that the first report held out more encouragement to job selectors than the last. The conclusion was reached then that 'the final weighted results show a difference of 48 % in accident rate between those above and those below the average in the tests' (Farmer & Chambers, 1926, p. 36). The study of bus drivers led to the following conclusions (Farmer & Chambers, 1939, p. 36). If the worst quartile of testees had been eliminated the subsequent accident rate on the retained would be 7 % less than it actually was for the total. If not only the worst quartile, but any other drivers who had three or more accidents in the first year were eliminated, a reduction of 44 % of the exposed to risk, the improvement would have been 13 %. There is, of course, no inconsistency between the results. Dockyard apprentices and employees of a wholesale confectioner do not have to pass the rigorous tests endured by would-be bus drivers. The eliminating tests are unlike those of the field psychologists, but they may well overlap them. Farmer & Chambers give the test scores of 128 of the 166 bus drivers. I find that the thirty-eight drivers who were not tested did not have accident scores significantly different from those of the 128, so we may regard the latter as fairly representative.

I make the correlation between accident score and test score for the first and fifth years 0·12 in each case; neither is 'significant' by itself, but, even assuming that there is really a positive correlation, its order of magnitude is too small to have much practical value if the regression is linear. Naturally what the field workers would like would be an unselected population of future motor-car drivers, all of whom are tested and all of whom are obliged to report every accident, however trivial. Outside a police state such a statisticians' paradise is a dream—or nightmare.

REFERENCES

CHAMBERS, E. G. (1941). *J.R. Statist. Soc. Suppl.* **7**, 89, 95.
FARMER, E. & CHAMBERS, E. G. (1926). *Rep. Industr. Fatig. Res. Bd*, no. 38. H.M.S.O.
FARMER, E. & CHAMBERS, E. G. (1939). *Rep. Industr. Hlth Res. Bd*, no. 84.
GREENWOOD, M. (1941). *J.R. Statist. Soc. Suppl.* **7**, 107.
GREENWOOD, M. & YULE, G. U. (1920). *J.R. Statist. Soc.* **83**, 255.
IRWIN, J. O. (1941). *J.R. Statist. Soc. Suppl.* **7**, 101.
NEWBOLD, E. M. (1926). *Rep. Industr. Fatig. Res. Bd*, no. 34. H.M.S.O.
NEWBOLD, E. M. (1927). *J.R. Statist. Soc.* **83**, 255.
PEARSON, K. (1917). *Biometrika*, **11**, 139.
SMITH, MAY (1943). *An Introduction to Industrial Psychology.* London.
YULE, G. U. (1941). *J.R. Statist. Soc. Suppl.* **7**, 91.

ORGANIC CORRELATION AND ALLOMETRY

By K. A. KERMACK and J. B. S. HALDANE, *University College, London*

In studies on organic correlation it is frequently desired to fit a straight line to show the general trend of the relationship between two variates. The residual scatter of the observed points about any such line is often due mainly to the biological variability of the material; and the errors of measurement may only account for a very small part of it. For such a case the conventionally used regression lines are quite unsuitable, since here the terms 'dependent variate' and 'independent variate' have no real meaning. It would perhaps be more reasonable in our case to follow Karl Pearson (1901) and choose the line which minimizes, not the sum of the squares of the deviations of one of the variates as do the regression lines, but rather the sum of squares of the normal deviations of the observed points. This line is the major axis of the correlation surface, making an angle θ with the x axis, where

$$\tan 2\theta = \frac{2\mu_{11}}{\mu_{20}-\mu_{02}}, \tag{1 A}$$

$$\tan\theta = -\frac{(\mu_{20}-\mu_{02})+\sqrt{[4\mu_{11}^2+(\mu_{20}-\mu_{02})^2]}}{2\mu_{11}}, \tag{1 B}$$

and was first suggested by Adcock (1878) from the rather different point of view of finding the best line when the errors of measurement of both variates are equal. For us it suffers from the disadvantage that, while invariant under rotation of the axes, it is not invariant under change of scale. In practice it is the latter that is important, invariance under rotation being normally of negligible advantage. As suggested by Jones (1937) and Teissier (1948), a line invariant under change of scale may be obtained by considering the major axis of the distribution of two variates expressed in standard measure. Then, since in this measure $\mu_{20} = \mu_{02} = 1$ by definition, $\tan\theta$ is also identically equal to unity, irrespective of the value of μ_{11}.

Therefore, the equation to the line in terms of the original variables is

$$\frac{y-\bar{y}}{\sigma_y} = \frac{x-\bar{x}}{\sigma_x},$$

and so

$$y = \frac{\sigma_y}{\sigma_x}x + \left(\bar{y} - \frac{\sigma_y}{\sigma_x}\bar{x}\right). \tag{2}$$

This line may be called the reduced major axis. It may be seen that, not only is it invariant under change of scale, but, in addition, its computation involves only the sums and the sums of squares of the two variates, the computation of the sum of products not being involved. A further advantage will be seen by a consideration of the standard error of the estimate of σ_y/σ_x in equation (2) compared with the standard error of the estimate of $\tan\theta$, which is involved if the major axis is fitted. These may all be obtained up to the order of n^{-1} by straightforward methods, as below.

As in equation (1), but considering here the sample, with T the estimate of θ, we have

$$\tan 2T = \frac{2m_{11}}{m_{20}-m_{02}}.$$

Now consider the standard error; we proceed as follows:

$$\frac{\delta(\tan 2T)}{\tan 2\theta} = \frac{\delta m_{11}}{\mu_{11}} - \frac{\delta m_{20}}{\mu_{20}-\mu_{02}} + \frac{\delta m_{02}}{\mu_{20}-\mu_{02}}.$$

Squaring and summing gives

$$\operatorname{var}(\tan 2T) = \frac{4}{n(\mu_{20}-\mu_{02})^2}\left[\mu_{22} + \mu_{11}^2\frac{(\mu_{40}+\mu_{04}-2\mu_{22})}{(\mu_{20}-\mu_{02})^2} + 2\mu_{11}\frac{(\mu_{13}-\mu_{31})}{\mu_{20}-\mu_{02}}\right], \tag{3}$$

so that $$\operatorname{var} T = \frac{\mu_{22}(\mu_{20}-\mu_{02})^2 + \mu_{11}^2(\mu_{40}+\mu_{04}-2\mu_{22}) + 2\mu_{11}(\mu_{13}-\mu_{31})(\mu_{20}-\mu_{02})}{n[(\mu_{20}-\mu_{02})^2+4\mu_{11}^2]^2}, \tag{4}$$

and so $$\operatorname{var}(\tan T) = \frac{(\mu_{20}-\mu_{02})^4}{4\mu_{11}^4}\left[\frac{4\mu_{11}^2}{(\mu_{20}-\mu_{02})^2} + 1 \pm \sqrt{\left(1+\frac{4\mu_{11}^2}{(\mu_{20}-\mu_{02})^2}\right)}\right]^2 \operatorname{var} T. \tag{5}$$

$\sqrt{\left(1+\frac{4\mu_{11}^2}{(\mu_{20}-\mu_{02})^2}\right)}$ takes the negative sign if $\tan 2\theta$ is positive.

In the normal case (3), (4) and (5) reduce to

$$\operatorname{var}(\tan 2T) = \frac{(1-\rho^2)(1+\tan^2 2\theta)\tan^2 2\theta}{n}, \tag{6}$$

$$\operatorname{var} T = \frac{(1-\rho^2)\sin^2 2\theta}{4\rho^2 n}, \tag{7}$$

and $$\operatorname{var}(\tan T) = \frac{(1-\rho^2)\tan^2\theta}{n\rho^2}, \tag{8}$$

$$\sigma_{\tan T} = \left(\frac{1-\rho^2}{n}\right)^{\frac{1}{2}}\frac{\tan\theta}{\rho}. \tag{9}$$

Equation (8) in certain cases takes simpler forms. If $\rho = 0$

$$\begin{aligned} \operatorname{var}(\tan T) &= \frac{\sigma_x^2-\sigma_y^2}{n(\sigma_x^2-\sigma_y^2)^2}, && \text{if} \quad \sigma_x^2 > \sigma_y^2 \\ &= \infty, && \text{if} \quad \sigma_x^2 < \sigma_y^2, \end{aligned}$$

and if $$\sigma_x^2 = \sigma_y^2,$$

$$\operatorname{var}(\tan T) = \frac{1-\rho^2}{n\rho^2}.$$

For the slope of the reduced major axis of equation (2), the sampling variance of s_y/s_x is similarly obtained:

$$\frac{\delta(m_{02}/m_{20})}{\mu_{02}/\mu_{20}} = \frac{\delta m_{02}}{\mu_{02}} - \frac{\delta m_{20}}{\mu_{20}}.$$

Squaring and summing $$\operatorname{var}\left(\frac{m_{02}}{m_{20}}\right) = \frac{\mu_{02}^2}{n\mu_{20}^2}\left(\frac{\mu_{40}}{\mu_{20}^2} + \frac{\mu_{04}}{\mu_{02}^2} - \frac{2\mu_{22}}{\mu_{20}\mu_{02}}\right),$$

and $$\operatorname{var}\left(\frac{s_y}{s_x}\right) = \frac{\mu_{02}}{4n\mu_{20}}\left(\frac{\mu_{40}}{\mu_{20}^2} + \frac{\mu_{04}}{\mu_{02}^2} - \frac{2\mu_{22}}{\mu_{20}\mu_{02}}\right). \tag{10}$$

In the normal case, equation (10) becomes

$$\operatorname{var}\left(\frac{s_y}{s_x}\right) = \frac{\sigma_y^2}{\sigma_x^2}\frac{(1-\rho^2)}{n}. \tag{11}$$

Equation (11) is identical with that obtained for the variance of the regression coefficient in a bivariate normal population—a rather surprising result, due to Teissier (1948).

Table 1 gives the efficiency of $\tan T$ in terms of s_y/s_x, as a percentage, in a number of cases. It can be seen from this that the use of the statistic derived from the major axis, rather than the other, always leads to a loss of efficiency, which can, if σ_x^2 and σ_y^2 are nearly equal, and the coefficient of correlation small, be enormous.

Table 1. *Comparison of the accuracy with which the slopes of the reduced major axis and the major axis can be estimated*

$$\frac{\operatorname{var}(s_y/s_x)}{\operatorname{var}(\tan T)} \times 100.$$

ρ \ σ_y/σ_x	0	0·1	0·4	0·7	1·0
0·0	100·0	98·0	70·6	26·0	0·0
0·1	100·0	98·0	70·9	27·0	1·0
0·5	100·0	98·5	78·4	47·3	25·0
0·9	100·0	99·6	94·7	87·4	81·0
1·0	100·0	100·0	100·0	100·0	100·0

The exact distribution of the ratio s_y/s_x is known in the bivariate normal case, and is given by Bose (1935). It is

$$dF = \frac{2(1-\rho^2)^{\frac{1}{2}(n-1)}}{B\left(\frac{n-1}{2}, \frac{n-1}{2}\right)} \frac{\omega^{n-2}}{(1+\omega^2)^{n-1}} \left\{1 - \frac{4\rho^2\omega^2}{(1+\omega^2)^2}\right\}^{-\frac{1}{2}n}; \tag{12}$$

where $\omega = \frac{s_y \sigma_x}{s_x \sigma_y}$. Finney (1938) should be consulted for practical details as to the use of this distribution. The distribution of $\tan T$ is unknown.

The sampling variance of $\bar{y} - \bar{x}\tan T$ may be obtained as below. For a normal population

$$\delta(\bar{y} - \bar{x}\tan T) = \delta\bar{y} - \tan\theta\,\delta\bar{x} - \mu_{10}\,\delta\tan T,$$

so
$$\operatorname{var}(\bar{y} - \bar{x}\tan T) = \frac{1}{n}\left[(\sigma_y - \sigma_x\tan\theta)^2 + (1-\rho)\tan\theta\left(2\sigma_x\sigma_y + \frac{\mu_{10}^2\tan\theta(1+\rho)}{\rho^2}\right)\right]. \tag{13}$$

The sampling variance of $(\bar{y} - \bar{x}s_y/s_x)$ may also be easily obtained as below:

$$\delta(\bar{y} - \bar{x}s_y/s_x) = \delta\bar{y} - \delta\bar{x}\,\sigma_y/\sigma_x - \mu_{10}\,\delta(s_y/s_x),$$

so, for a normal population,

$$\operatorname{var}(\bar{y} - \bar{x}s_y/s_x) = \frac{\sigma_y^2}{n}(1-\rho)\left(2 + \frac{\mu_{10}^2(1+\rho)}{\sigma_x^2}\right), \tag{14}$$

a very simple result.

ALLOMETRIC CORRELATION*

In many cases of organic correlation it is found that the distribution of the variates is heteroscedastic and skew. Often, however, the distribution of the logarithms is homoscedastic and more nearly normal. Hemmingsen (1934) discusses this point at some length, and gives many examples. In addition, it may be found that the fitted straight line does not pass through the origin. This may be due to a significant curvature in the trend of the two variates, and it may be more useful to fit a line of the form $y = \beta x^{\alpha}$ (allometric growth of Huxley (1932)). In almost all cases where this has been done in the past, one of the regression lines has been fitted to the logarithms of the data. For the reasons given above, it is suggested that considerable advantages are obtained in many cases if the reduced major axis of the logarithms is taken. This may be done in a very simple way if the distribution of the logarithms can be assumed to be normal.

---

* In the following sections of the paper it has not always been possible to distinguish between the population parameters and their estimates obtained from the sample, by the use, in the usual way, of two separate sets of symbols. It is not anticipated, however, that any great inconvenience will thereby be caused to the reader, the meaning of the symbol being obvious in each case.

The numerical examples given below will show that the accuracy of the results obtained is not very sensitive to deviations from normality.

Let $x = e^X$, $y = e^Y$, where X and Y are normally correlated; and let $v_x = \sigma_x/\bar{x}$ and $v_y = \sigma_y/\bar{y}$, in other words let $100v_x$ and $100v_y$ be the coefficients of variation. Then if

$$\exp[\bar{X}t + \tfrac{1}{2}\sigma_X^2 t^2] = M(t) = \text{moment generating function of } x,$$

$$\overline{x^r} = \overline{e^{rX}} = M(r),$$

and so

$$\bar{x} = \exp[\bar{X} + \tfrac{1}{2}\sigma_X^2], \tag{15}$$

$$\overline{x^2} = \exp[2\bar{X} + 2\sigma_X^2]. \tag{16}$$

Thus

$$\exp[\sigma_X^2] = 1 + v_x^2$$

and

$$\sigma_X^2 = \log(1 + v_x^2), \tag{17}$$

$$\tau_Y^2 = \log(1 + v_y^2). \tag{18}$$

Thus we see that

$$\frac{\sigma_X}{\sigma_Y} = \left[\frac{\log(1 + v_y^2)}{\log(1 + v_x^2)}\right]^{\frac{1}{2}} = \alpha, \tag{19}$$

if we take α as the slope of the reduced major axis of the X, Y distribution. As (19) is the ratio of two logarithms, either natural or decimal logarithms may be used in computation.

It may readily be seen that the accuracy of equation (19) will not be very sensitive to a moderate deviation from normality of the distribution of the logarithms.

Let $\bar{\zeta} = \bar{\eta} = 0$, $\overline{\zeta^2} = 1 + \rho$, $\overline{\eta^2} = 1 - \rho$, and let $\kappa_3, \kappa_4, \ldots$ be the other cumulants of the distribution of ζ, κ_3', κ_4', ... those of η. Let ζ and η be independent. Now let

$$X = m + \frac{\sigma}{\sqrt{2}}(\zeta + \eta),$$

$$Y = \alpha m + \log\beta + \frac{\alpha\sigma}{\sqrt{2}}(\zeta - \eta).$$

Then

$$\operatorname{var} X = \sigma^2, \quad \operatorname{var} Y = \alpha^2\sigma^2$$

and

$$S(X - \bar{X})(Y - \bar{Y}) = \alpha\rho\sigma^2.$$

So X and Y have correlation ρ, regardless of the higher cumulants of ζ and η. Then the cumulant generating function of X is

$$\kappa_X(t) = mt + \tfrac{1}{2}\sigma^2 t^2 + \mathop{S}_{r=3}^{r=\infty}\left(\frac{\sigma}{\sqrt{2}}\right)^r (\kappa_r + \kappa_r')\frac{t^r}{r!},$$

and that of Y

$$\kappa_Y(t) = (\alpha m + \log\beta)t + \tfrac{1}{2}\alpha^2\sigma^2 t^2 + \mathop{S}_{r=3}^{r=\infty}\left(\frac{\alpha\sigma}{\sqrt{2}}\right)^r (\kappa_r - \kappa_r')\frac{t^r}{r!},$$

and let

$$x = e^X, \quad y = e^Y.$$

Then as before

$$\overline{x^r} = \overline{e^{rX}} = e^{\kappa(r)},$$

so that
$$\bar{x} = \exp[m+\tfrac{1}{2}\sigma^2]\exp\left[\mathop{S}_{r=3}^{r=\infty}\left(\frac{\sigma}{\sqrt{2}}\right)^r(\kappa_r+\kappa_r')\frac{1}{r!}\right],$$
and
$$\overline{x^2} = \exp[2m+2\sigma^2]\exp\left[\mathop{S}_{r=3}^{r=\infty}(\sqrt{2}\,\sigma)^r(\kappa_r+\kappa_r')\frac{1}{r!}\right].$$
Clearly, then
$$1+v_x^2 = \exp\left[\sigma^2+\mathop{S}_{r=3}^{r=\infty}\left(\frac{\kappa_r+\kappa_r'}{r!}\right)(2^{r-1}-1)\frac{\sigma^r}{\sqrt{(2)^{r-2}}}\right],$$
and
$$\log(1+v_x^2) = \sigma^2+\mathop{S}_{r=3}^{r=\infty}\left(\frac{\kappa_r+\kappa_r'}{r!}\right)(2^{r-1}-1)\frac{\sigma^r}{\sqrt{(2)^{r-2}}}.$$
Similarly
$$\log(1+v_y^2) = \alpha^2\sigma^2+\mathop{S}_{r=3}^{r=\infty}\left(\frac{\kappa_r-\kappa_r'}{r!}\right)(2^{r-1}-1)\frac{\alpha^r\sigma^r}{\sqrt{(2)^{r-2}}},$$
giving
$$\frac{\log(1+v_y^2)}{\log(1+v_x^2)} = \alpha^2\left[1+\frac{(\alpha-1)\kappa_3-(\alpha+1)\kappa_3'}{2\sqrt{2}}.\sigma+\frac{\sigma^2}{48}\{7(\alpha^2-1)\kappa_4-7(\alpha^2+1)\kappa_4' \right.$$
$$\left. -6(\alpha-1)\kappa_3(\kappa_3+\kappa_3')+6(\alpha+1)\kappa_3'(\kappa_3+\kappa_3')\}+\dots\right].$$
Putting $\kappa_3 = (1+\rho)^{\frac{3}{2}}\gamma_1$, $\kappa_4 = (1+\rho)^2\gamma_2$, $\kappa_3' = (1-\rho)^{\frac{3}{2}}\gamma_1'$, $\kappa_4' = (1-\rho)^2\gamma_2'$,
and so on, we have
$$\frac{\log(1+v_y^2)}{\log(1+v_x^2)} = \alpha^2\left[1+\frac{(\alpha-1)(1+\rho)^{\frac{3}{2}}\gamma_1-(\alpha+1)(1-\rho)^{\frac{3}{2}}\gamma_1'}{2\sqrt{2}}\sigma \right.$$
$$+\frac{\sigma^2}{48}\{7(\alpha^2-1)(1+\rho)^2\gamma_2-7(\alpha^2+1)(1-\rho)^2\gamma_2'-6(\alpha-1)(1+\rho)^3\gamma_1^2$$
$$\left. +6(\alpha+1)(1-\rho)^3\gamma_1'^2+12(1-\rho^2)^{\frac{3}{2}}\gamma_1\gamma_1'\}+O(\sigma^3)\right].$$

Supposing γ_1, γ_2, γ_1', γ_2', etc., to be quantities of the order of unity, the first term of the correction factor is small with $(\alpha-1)\sigma$ or $(1-\rho)^{\frac{3}{2}}\sigma$. σ is approximately the coefficient of variation, and will rarely exceed 0·2, while $\alpha-1$ and $1-\rho$ are generally of the order of 0·1 or 0·2. Thus α^2 will rarely be out by more than a factor of 4 % or α by a factor of more than 2 %. The most serious errors are likely to occur when α is fairly large, and ρ small.

In addition, we see from equation (19) that if two variates with normally correlated logarithms have different coefficients of variation, then either the axis is curved (i.e. $\alpha \neq 1$) and allometry must be suspected, or it does not pass through the origin. In biology there are a number of examples of the latter, amongst organs which do not appear until the animal has reached a certain minimum size, antler length in male deer being a case in point. In most cases, however, for two organs, both present in the embryo, the axis must pass through the origin.

Further, from (15) and (16), it can be seen that
$$\overline{e^X} = \bar{x}(1+v_x^2)^{-\frac{1}{2}},$$
and
$$\bar{X} = \log\bar{x}-\tfrac{1}{2}\log(1+v_x^2), \tag{20}$$
$$\bar{Y} = \log\bar{y}-\tfrac{1}{2}\log(1+v_y^2). \tag{21}$$
Thus
$$\bar{Y}-\frac{\sigma_Y}{\sigma_X}\bar{X} = \log\bar{y}-\tfrac{1}{2}\log(1+v_y^2)-\left[\frac{\log(1+v_y^2)}{\log(1+v_x^2)}\right]^{\frac{1}{2}}[\log\bar{x}-\tfrac{1}{2}\log(1+v_x^2)], \tag{22}$$
(22) giving our estimate of $\log\beta$ by the insertion of the sample values. It should be noted that while m_1' and m_2 are unbiased estimates of μ_1' and μ_2, $\log(1+m_2/m_1'^2)$ is not an unbiased

estimate of $\log(1+\mu_2/\mu_1'^2)$, nor is $\log m_1' - \frac{1}{2}\log(1+m_2/m_1'^2)$ of $\log\mu_1' - \frac{1}{2}\log(1+\mu_2/\mu_1'^2)$. The unbiased estimates have been derived in a separate paper*. They are, to an order n^{-1},

$$\log\left(1+\frac{\mu_2}{\mu_1'^2}\right) = \log\left(1+\frac{m_2}{m_1'^2}\right) - \frac{4m_2^2}{nm_1'^4},$$

and
$$\log\mu_1' - \tfrac{1}{2}\log\left(1+\frac{\mu_2^2}{\mu_1'^2}\right) = \log m_1' - \tfrac{1}{2}\log\left(1+\frac{m_2}{m_1'^2}\right) - \frac{m_2}{2m_1'^2}.$$

The correction to α is therefore

$$\left(\frac{\log(1+\mu_{20}/\mu_{10}'^2)}{\log(1+\mu_{02}/\mu_{01}'^2)}\right)^{\frac{1}{2}} = \left(\frac{\log(1+m_{20}/m_{10}'^2)}{\log(1+m_{02}/m_{01}'^2)}\right)\left(1-\frac{2m_{02}}{nm_{01}'^2}(\alpha^2-1)\right).$$

It can be seen that all corrections are normally quite negligible. The transformation of a grouped variate will introduce further bias, which again will be usually negligible. The reference quoted should be consulted for further details if required. These corrections have not been employed in computing the examples given in the present paper.

The sampling variances of the estimates obtained from (19) and (22) can be obtained by substitution of the sample values in equations (11) and (14) above. The variance of $\exp(\overline{Y} - \overline{X}\, s_Y/s_X)$, the estimate of β, may then be obtained at once by remembering that

$$\operatorname{var}(e^x) = e^{2\bar{x}} \operatorname{var} x. \tag{23}$$

The exact distribution given in equation (12) may also be employed. It is clear that in any case the allometry is significant if α is significantly different from unity. The method given by Finney (1938) is useful for testing in this case. Similarly, the coefficient of correlation ρ of X and Y may be obtained in terms of the correlation coefficient ρ' of x and y,

$$\begin{aligned}\overline{xy} &= \overline{\exp[X+Y]} \\ &= \exp[\overline{X}+\overline{Y}+\tfrac{1}{2}(\sigma_X^2+\sigma_Y^2+2\rho\sigma_X\sigma_Y)] \\ &= \bar{x}\,\bar{y}\exp\{\rho[\log(1+v_x^2)\log(1+v_y^2)]^{\frac{1}{2}}\}.\end{aligned}$$

Therefore
$$\rho = \frac{\log(1+\rho' v_x v_y)}{[\log(1+v_x^2)\log(1+v_y^2)]^{\frac{1}{2}}}. \tag{24}$$

And since
$$\mu_{11} = \rho\sigma_X\sigma_Y = \log(1+\rho' v_x v_y), \tag{25}$$

the regression equations of the logarithmic data may be obtained by substitution in the well-known formula

$$y = \frac{\mu_{11}}{\mu_{02}}x - \left(\bar{y} - \frac{\mu_{11}}{\mu_{02}}\bar{x}\right). \tag{26}$$

Similarly, the major axis of the logarithmic data may be obtained if desired from equation (1), and the variance of $\tan T$ from equation (8).

The sampling variances of equation (26) may be obtained by substitution of the sample values in the well-known results

$$\operatorname{var}\left(\frac{m_{11}}{m_{02}}\right) = \frac{\mu_{20}}{\mu_{02}}\frac{(1-\rho^2)}{n}, \tag{27}$$

$$\operatorname{var}\left(\bar{y} - \frac{m_{11}}{m_{02}}\bar{x}\right) = \frac{\mu_{20}}{n}(1-\rho^2)\left(1+\frac{\mu_{10}^2}{\mu_{02}}\right). \tag{28}$$

---

* It is hoped to publish this paper shortly in *Biometrika*.

A SHORTER METHOD

If v_x and v_y become small, so that v^2 and higher terms of v may be neglected, then from equation (19)

$$\alpha = \frac{\sigma_Y}{\sigma_X} = \frac{v_y}{v_x}, \tag{29}$$

while in the same way

$$\rho\frac{\sigma_Y}{\sigma_X} = \rho'\frac{v_y}{v_x}. \tag{30}$$

These results may be obtained in another way.

If the correlation between the two variates be high and positive, and their relation approximately linear, then the appropriate straight line may be taken as a tangent to the allometry curve at the means of the variates. Then since

$$y = \beta x^{\alpha},$$

and

$$\frac{dy}{dx} = \alpha\beta x^{\alpha-1},$$

$$\alpha = \frac{\bar{x}}{\bar{y}}\frac{dy}{dx}.$$

Thus for the reduced major axis

$$\alpha = \frac{\bar{x}}{\bar{y}}\frac{\sigma_y}{\sigma_x} = \frac{v_y}{v_x},$$

as in equation (29). For the regression lines

$$\alpha = \frac{\bar{x}}{\bar{y}}\rho'\frac{\sigma_y}{\sigma_x} = \rho'\frac{v_y}{v_x},$$

as in equation (30).

This method does not apply to the major axis, since

$$-\frac{(v_x^2-v_y^2)+\sqrt{[(4\rho'^2v_x^2v_y^2+(v_x^2-v_y^2)^2]}}{2v_xv_y\rho'} \neq \frac{\bar{x}}{\bar{y}}\frac{\{-(\sigma_X^2-\sigma_Y^2)+\sqrt{[4\rho^2\sigma_X\sigma_Y+(\sigma_X^2-\sigma_Y^2)^2]}\}}{2\sigma_X\sigma_Y\rho},$$

and the use of the short method is not recommended in this case. In all cases

$$\log\beta = \log\bar{y}-\alpha\log\bar{x}. \tag{31}$$

The sampling variance of the estimates of

$$\frac{v_y}{v_x},\quad \log\bar{y}-\frac{v_y}{v_x}\log\bar{x},\quad \rho\frac{v_y}{v_x}\quad\text{and}\quad \log\bar{y}-\rho\frac{v_y}{v_x}\log\bar{x}$$

may be obtained quite simply as below. For the reduced major axis

$$\frac{\delta(v_y^2/v_x^2)}{v_y^2/v_x^2} \doteq \frac{\delta m_{20}}{m_{20}}-\frac{\delta m_{02}}{m_{02}}-\frac{2\delta m'_{10}}{m_{10}}+\frac{2\delta m'_{01}}{m'_{01}}.$$

Squaring and summing for a normal parent population

$$\operatorname{var}\left(\frac{v_y}{v_x}\right) = \frac{1}{n}\frac{v_y^2}{v_x^2}(1-\rho'^2+v_x^2+v_y^2-2\rho'v_xv_y), \tag{32}$$

While $\delta\left(\log\bar{y}-\frac{v_y}{v_x}\log\bar{x}\right) = \frac{1}{\bar{y}}\left[\delta\bar{y}\left(1+\frac{v_y}{v_x}\log\bar{x}\right)-\delta\bar{x}(1+\log\bar{x})\frac{s_y}{s_x}-\delta\left(\frac{s_y}{s_x}\right)\bar{x}\log\bar{x}\right],$
so that

$$\operatorname{var}\left(\log\bar{y}-\frac{v_y}{v_x}\log\bar{x}\right)$$
$$= \frac{v_y^2}{n}\left[\log^2\bar{x}\left(1-\frac{v_y}{v_x}\right)^2+(1-\rho')\left\{2(1+\log\bar{x})\left(1+\frac{v_y}{v_x}\log\bar{x}\right)+(1+\rho')\frac{\log^2\bar{x}}{v_x^2}\right\}\right]. \tag{33}$$

For the regression lines $\dfrac{\delta(\rho' v_y/v_x)}{\rho' v_y/v_x} = \dfrac{\delta m_{11}}{m_{11}} - \dfrac{\delta m_{02}}{m_{02}} + \dfrac{\delta m'_{01}}{m'_{01}} - \dfrac{\delta m'_{10}}{m'_{10}}$,
and, squaring and summing, as before, for a normal population

$$\operatorname{var}\left(\rho'\frac{v_y}{v_x}\right) = \frac{1}{n}\frac{v_y^2}{v_x^2}[1-\rho'^2(1-v_x^2-v_y^2+2\rho' v_x v_y)]. \tag{34}$$

Further, we see that

$$\delta\left(\log\bar{y}-\rho'\frac{v_y}{v_x}\log\bar{x}\right) = \frac{1}{\bar{y}}\left[\delta\bar{y}\left(1+\frac{v_y}{v_x}\log\bar{x}\right)-\delta\bar{x}\rho'\frac{v_y}{v_x}(\log\bar{x}+1)-\delta\left(\rho'\frac{v_y}{v_x}\right)\bar{x}\log\bar{x}\right];$$

squaring and summing

$$\operatorname{var}\left(\log\bar{y}-\rho'\frac{v_y}{v_x}\log\bar{x}\right) = \frac{v_y^2}{n}\left[(1-\rho'^2)\left(1+\frac{\log^2\bar{x}}{v_x^2}+2\rho'\frac{v_y}{v_x}\log\bar{x}\right)+\rho'^2\left(1+\frac{v_y}{v_x}(1-2\rho')\right)\right]. \tag{35}$$

In all cases, as in equation (23), $\operatorname{var} b = b^2 \operatorname{var}(\log b)$.

It can be seen that equations (32), (33), (34) and (35) only reduce to those obtained in the exact case under the following conditions: the coefficients of variation should be negligible compared to unity and subequal, and, for equations (33) and (35), the coefficient of correlation should be large and positive. Under any other conditions the short method may give misleading results.

Below, in Appendices, are given an Example and a summary of the main formulae developed.

APPENDICES

I. *Illustrative example*

An example will be given. It will be taken from a population of 338 specimens of *Micraster cor-anguinum* from the chalk of Northfleet, Kent. On these specimens the total length (l), and the height to the apical system (h) were measured. Table 2 gives these results arranged as a correlation table. From an examination of this table it may at once be seen that the distribution is both negatively skew and heteroscedastic, while the trend of the observations shows a slight but distinct curvature. From Table 3 it can be seen that the correlation between the two variates is high (0·8848), and that the fitted straight lines do not pass through the origin. Table 4 shows the data replotted as natural logarithms; the distribution now seems to be reasonably homoscedastic, although the negative skewness is, of course, increased. Table 5 gives certain relevant parameters for this distribution. In Table 6 the parameters

Table 3. Micraster cor-anguinum. *Parent population*

Apical height (h)	Total length (l)
Mean ($\bar{h}$) = 30·903 mm. Variance (s_h^2) = 21·93 sq.mm. Coefficient of variation × 0·01(v_h) = 0·1515	Mean (l) = 50·078 mm. Variance (s_l^2) = 92·87 sq.mm. Coefficient of variation × 0·01(v_l) = 0·1924

Coefficient of correlation (r) = 0·8848 ± 0.0118

s_h/s_l = 0·4860 ± 0·0123, $\bar{h}-(s_h/s_l)\,l$ = 6·567 ± 0·623 mm.,
$r(s_h/s_l)$ = 0·4300 ± 0·0123, $\bar{h}-r(s_h/s_l)\,l$ = 9·370 ± 0·632 mm.,
$\tan T$ = 0·4494 ± 0·0129, $\bar{h}-\tan T l$ = 8·398 ± 0·656 mm.

Table 2. Micraster cor-anguinum. *Table of length × apical height*

Length in mm. pical eight n mm.	23·995–24·995	24·995–	25·995–	26 995–	27·995–	28·995–	29·995–	30 995–	31·995–	32·995–	33·995–	34·995–	35·995–	36·995–	37·995–	38·995–	39·995–	40·995–	41·995–	42·995–	43·995–	44·995–	45·995–	46·995–	47·995–	48 995–	49·995–	50 995–	51·995–	52·995–	53·995–	54·995–	55 995–	56·995–	57 995–	58·995–	59·995–	60·995–	61·995–	62·995–	63·995–	64·995–	65·995–	66·995–	67·995–	68·995–	69 995–	70·995–71·995	Total
·995–41·995	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	1
995–	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	1
995–	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	1	.	1	.	1	1	.	.	.	.	.	.	5
995–	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	1	.	1	.	.	.	.	.	.	1	1	5
995–	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	1	.	1	2	2	2	.	.	.	.	.	.	.	.	.	.	9
995–	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	1	1	1	.	2	3	2	.	1	2	2	1	.	.	.	.	.	.	17
995–	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	.	1	2	3	2	1	2	2	1	1	1	2	.	.	1	.	.	.	.	.	21
995–	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	1	.	3	3	1	1	3	1	1	2	6	3	1	1	3	1	1	.	.	.	.	.	.	33
995–	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	3	1	1	1	3	3	1	3	3	8	3	4	1	.	.	.	.	2	1	.	.	.	.	.	38
995–	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	2	3	3	4	2	4	4	.	3	1	2	.	.	.	.	.	.	.	.	.	.	.	.	.	29
995–	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	2	4	2	1	4	1	2	4	2	1	1	1	–	2	2	1	.	.	.	.	.	1	.	.	.	.	.	.	32
995–	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	2	2	.	1	2	2	4	.	2	.	.	.	1	.	1	.	1	.	.	.	.	.	.	.	.	.	.	19
995–	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	1	2	5	2	1	1	.	2	2	1	1	3	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	24
995–	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	2	1	1	1	4	3	.	2	1	.	1	.	.	.	1	1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	20
995–	.	.	.	.	.	.	.	.	.	.	.	.	1	.	3	.	1	1	3	3	1	1	1	1	1	1	.	.	.	1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	20
995–	.	.	.	.	.	.	.	.	.	.	.	.	.	.	3	.	2	.	2	.	.	1	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	9
995–	.	.	.	.	.	.	.	.	.	.	.	4	4	1	3	1	.	.	1	.	1	.	.	.	1	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	17
995–	.	.	.	.	.	.	.	.	.	2	.	1	3	2	1	.	.	.	1	.	.	.	1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	12
995–	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
95–	.	.	.	.	.	1	.	3	1	.	1	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	8
95–	.	.	.	.	.	2	2	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5
95–	.	.	.	.	.	1	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2
95–	.	1	.	.	1	.	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5
95–	.	.	1	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2
95–	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
95–	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
95–15·995	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
Total	1	1	2	.	2	4	5	6	1	3	1	7	8	3	11	3	6	2	10	9	13	11	9	5	13	12	14	18	11	17	12	12	12	14	14	20	11	6	4	9	6	6	2	.	.	.	1	1	338

Table 4. Micraster cor-anguinum. *Correlation table formed of natural logarithms of apical height (H) and length (L)*

H \ L	3·25	3·35	3·45	3·55	3·65	3·75	3·85	3·95	4·05	4·15	4·25	Total
3·75	—	—	—	—	—	—	—	—	—	1	—	1
3·65	—	—	—	—	—	—	—	2	10	14	2	28
3·55	—	—	—	—	—	—	3	23	46	22	1	95
3·45	—	—	—	—	—	3	25	37	21	2	—	88
3·35	—	—	—	1	8	16	19	12	2	1	—	59
3·25	—	—	—	10	11	7	3	1	1	—	—	33
3·15	—	1	2	7	3	1	1	—	—	—	—	15
3·05	—	3	6	—	—	—	—	—	—	—	—	9
2·95	2	2	3	—	—	—	—	—	—	—	—	7
2·85	—	—	1	—	—	—	—	—	—	—	—	1
2·75	2	—	—	—	—	—	—	—	—	—	—	2
Total	4	6	12	18	22	27	51	75	80	40	3	338

Table 5. *Logarithm population*

Logs of apical heights (H)	Logs of total length (L)
Mean ($\bar{H}$) $= 3·4183$ Variance (s_H^2) $= 0·02852$	Mean (L) $= 3·8914$ Variance (s_L^2) $= 0·04475$

Coefficient of correlation (r) $= 0·8865 \pm 0·0116$.

Table 6. Micraster cor-anguinum. $h = bl^a$

Method	a	b
Regression of log h on log l (calculated from logarithmic data)	$0·708 \pm 0·020$	$1·94 \pm 0·15$
Regression of log h on log l (assuming normally distributed logs)	$0·701 \pm 0·020$	$1·98 \pm 0·15$
Regression of log h on log l (short method)	$0·697 \pm 0·020$	$2·02 \pm 0·16$
Regression of log l on log h (calculated from logarithmic data)	$0·901 \pm 0·026$	0·918
Regression of log l on log h (assuming normally distributed logs)	$0·891 \pm 0·025$	0·944
Regression of log l on log h (short method)	$0·890 \pm 0·026$	0·950
Major axis (calculated from logarithmic data)	$0·776 \pm 0·022$	$1·49 \pm 0·13$
Major axis (assuming normally distributed logs)	$0·767 \pm 0·022$	$1·54 \pm 0·13$
Reduced major axis (calculated from logarithmic data)	$0·798 \pm 0·020$	$1·37 \pm 0·11$
Reduced major axis (assuming normally distributed logs)	$0·790 \pm 0·020$	$1·41 \pm 0·11$
Reduced major axis (short method)	$0·788 \pm 0·020$	$1·42 \pm 0·12$

of the equation of allometry ($y = \beta x^{\alpha}$) have been computed in three ways: from the logarithmic data of Table 4, from Table 2, in the one case on the assumption that the distribution of the logarithms is normal, and in the other assuming that the fitted straight line and the appropriate allometric curve are tangential at the means (short method). Results are given for the two regression lines, for the major axis, and for the reduced major axis. It can be seen that the agreement between the results obtained by the three different methods is very close, the results obtained by assuming the logarithms to be normally distributed being consistently better than those obtained by the short method. The marked negative skewness of the logarithmic distribution seems to have had singularly little effect on the accuracy of the results obtained by the method assuming normality. In addition, the value obtained for the coefficient of correlation by the latter method (0·8866 ± 0·0116) should be compared with that obtained directly from the logarithmic data (0·8865 ± 0·0116).

Incidentally these results show the very considerable disadvantage of using regression to calculate allometry. The value of α calculated from the regression of h on l is 0·71, from the regression of l on h 0·90, whereas 0·80 is a more satisfactory value.

II. *Summary of formulae*

1. *Fitting a straight line, $y = ax + b$.*

(1·1) Major axis:

$$a = \tan T\left(1 \pm \frac{1}{r'}\sqrt{\frac{1-r'^2}{n}}\right), \quad \tan 2T = \frac{2s_x s_y r'}{s_x^2 - s_y^2},$$

$$b = \bar{y} - \bar{x}\tan T \pm \sqrt{\left\{\frac{1}{n}(s_y - s_x \tan T)^2 + (1-r')\tan T\left[2s_x s_y + \frac{\bar{x}\tan T(1+r')}{r'^2}\right]\right\}}.$$

(1·2) Reduced major axis:

$$a = \frac{s_y}{s_x}\left(1 \pm \sqrt{\frac{1-r'^2}{n}}\right),$$

$$b = \bar{y} - \bar{x}a \pm s_y\sqrt{\left[\frac{(1-r')}{n}\left(2 + \bar{x}^2\frac{(1+r')}{s_x^2}\right)\right]}.$$

(1·3) Regression line:

$$a = r'\frac{s_y}{s_x} \pm \frac{s_y}{s_x}\sqrt{\frac{1-r'^2}{n}},$$

$$b = \bar{y} - \bar{x}a \pm s_y\sqrt{\left[\frac{(1-r'^2)}{n}\left(1 + \frac{\bar{x}^2}{s_x^2}\right)\right]}.$$

2. *Fitting allometry curve, $y = bx^a$, assuming normally distributed logarithms.*

(2·1) Major axis:

$$a = \tan T'\left(1 \pm \frac{1}{r}\sqrt{\frac{1-r^2}{n}}\right), \quad \tan 2T' = \frac{2\log(1+v_x v_y r')}{\log(1+v_x^2) - \log(1+v_y^2)},$$

$$b = \exp\{\log\bar{y} - \tfrac{1}{2}\log(1+v_y^2) - [\log\bar{x} - \tfrac{1}{2}\log(1+v_x^2)]\tan T'\}$$

$$\times\left(1 \pm \sqrt{\left\{\frac{I}{2}\left[(\log^{\frac{1}{2}}(1+v_y^2) - \log^{\frac{1}{2}}(1+v_x^2)\tan T')^2 + (1-r)\tan T'\left\{2\log^{\frac{1}{2}}(1+v_y^2)\log^{\frac{1}{2}}(1+v_x^2) + \frac{[\log\bar{x} - \tfrac{1}{2}\log(1+v_x^2)]\tan\theta'(1+r)}{r^2}\right\}\right]\right\}}\right).$$

(2·2) Reduced major axis:

$$a = \sqrt{\frac{\log(1+v_y^2)}{\log(1+v_x^2)}}\left(1 \pm \sqrt{\frac{1-r^2}{n}}\right),$$

$$b = \exp\{\log\bar{y} - \tfrac{1}{2}\log(1+v_y^2) - a[\log\bar{x} - \tfrac{1}{2}\log(1+v_x^2)]\}$$

$$\times\left(1 \pm a\sqrt{\left\{\frac{(1-r)}{n}(2\log(1+v_x^2) + (1+r)[\log x + \tfrac{1}{2}\log(1+v_x^2)])^2\right\}}\right).$$

(2·3) Regression line:

$$a = \frac{\log(1+r'v_xv_y)}{\log(1+v_x^2)} \pm \sqrt{\left(\frac{\log(1+v_y^2)}{\log(1+v_x^2)}\frac{(1-r^2)}{n}\right)},$$

$$b = (\exp\{\log\bar{y} - \tfrac{1}{2}\log(1+v_y^2) - a[\log\bar{x} - \tfrac{1}{2}\log(1+v_x^2)]\})$$

$$\times\left(1 \pm \sqrt{\left\{\frac{(1-r^2)}{n}\log(1+v_y^2)\left[1+\frac{(\log\bar{x}-\frac{1}{2}\log(1+v_x^2))^2}{\log(1+v_x^2)}\right]\right\}}\right).$$

(2·4) Coefficient of correlation:

$$r = \frac{\log(1+r'v_xv_y)}{\sqrt{[\log(1+v_x^2)\log(1+v_y^2)]}}.$$

3. *Short method.*

(3·2) Reduced major axis:

$$a = \frac{v_y}{v_x}\left(1 \pm \sqrt{\frac{1-r'^2+v_x^2+v_y^2-2r'v_xv_y}{n}}\right),$$

$$b = (\exp\{\log\bar{y} - a\log\bar{x}\})$$

$$\times\left(1 \pm \frac{v_y}{n^{\frac{1}{2}}}\sqrt{\left\{\log^2\bar{x}\left(1-\frac{v_y}{v_x}\right)^2 - (1-r')\left[2(1+\log\bar{x})\left(1+\frac{v_y}{v_x}\log\bar{x}\right)+\frac{(1+r')\log^2\bar{x}}{v_x^2}\right]\right\}}\right).$$

(3·3) Regression line:

$$a = r'\frac{v_y}{v_x}\left(1 \pm \sqrt{\frac{1-r'^2-1+v_x^2+v_y^2-2r'v_xv_y}{n}}\right),$$

$$b = (\exp\{\log\bar{y} - a\log\bar{x}\})\left[1 \pm v_y\sqrt{\left\{\frac{(1-r'^2)}{n}\left[1+\frac{\log^2\bar{x}}{v_x^2}+2r'\frac{v_y}{v_x}\log\bar{x}+r'^2\left(1+\frac{v_y}{v_x}(1-2r')\right)\right]\right\}}\right].$$

REFERENCES

ADCOCK, R. J. (1878). A problem in least squares. *Analyst*, **5**, 53.

BOSE, S. S. (1935). On the distribution of the ratio of variances of the two samples drawn from a given normal bivariate correlated population. *Sankhyā*, **2**, 65.

FINNEY, D. J. (1938). The distribution of the ratio of estimates of the two variances in a sample from a normal bivariate population. *Biometrika*, **30**, 190.

HEMMINGSEN, A. M. (1934). A statistical analysis of the differences in body size of related species. *Vidensk. Medd. naturh. Foren. Kbh.* **98**, 125.

HUXLEY, J. S. (1932). *Problems of Relative Growth.* London: Methuen.

JONES, H. E. (1937). Some geometrical considerations in the general theory of fitting lines and planes. *Metron*, **13**, no. 1, 21.

PEARSON, K. (1901). On lines and planes of closest fit to systems of points in space. *Phil. Mag.* (6), **2**, 559.

TEISSIER, G. (1948). La relation d'allometrie: sa signification statistique et Biologique. *Biometrics*, **4**, no. 1, 14.

THE PROBABILITY INTEGRAL TRANSFORMATION WHEN THE VARIABLE IS DISCONTINUOUS

BY F. N. DAVID AND N. L. JOHNSON

1. It is well known that given a continuous random variable, x, distributed with elementary probability law, $p(x)$, then the random variable

$$y = \int_{-\infty}^{x} p(x)\,dx$$

is distributed rectangularly in the interval $[0, 1]$. This is, however, only true for a *continuous* random variable x with known probability law. In a previous publication (1948) we have discussed the effect on the distribution of y if the parameters of the distribution of x are not known and have to be estimated from a random sample of n observations. In the present paper we discuss the moments of the distribution of the transformed variable y when x is discontinuous, first when the probability law of x is known and second when parameters of this law have to be estimated from observed data. In both these cases, if $p(x)$ be the elementary probability law of the discontinuous random variable, x, the transformed variable, which we shall call v, where

$$v = \sum_{-\infty}^{x} p(x),$$

will not be rectangularly distributed. The first of these cases has already been the subject of an arithmetical investigation by H. O. Lancaster (1949). Lancaster suggested a modification of Fisher's χ^2 test for the combination of tests of significance which has as its basis the assumption of rectangularity. This present investigation is carried out from a purely theoretical standpoint and in so far as the topics discussed here are common with those of Lancaster's paper, his empirical suggestions are confirmed.

2. We consider a random variable x which may take any one of a finite set of k mutually exclusive values, which are also the only possible, with probabilities $p_1, p_2, \ldots, p_k$, where

$$\sum_{j=1}^{k} p_j = 1.$$

The transformed variable, v, may therefore take any one of a set of values p_1, p_1+p_2, …, $p_1+p_2+\ldots+p_k$, with corresponding probabilities $p_1, p_2, \ldots, p_k$. From elementary considerations we have

$$\mathscr{E}(v) = \sum_{i=1}^{k} p_i \sum_{j=1}^{i} p_j = \sum_{i=1}^{k} p_i\left(1-\sum_{j=i+1}^{k} p_j\right) = \frac{1}{2}\left(1+\sum_{i=1}^{k} p_i^2\right).$$

The algebraic form of this expression suggests that we might redefine the variable, v. Thus instead of writing v as above we shall consider the variable u, where u is defined as

$$u_i = \sum_{j=1}^{i-1} p_j + \tfrac{1}{2}p_i \quad (i = 1, 2, \ldots, k),$$

and as before

$$P\{u = u_i\} = p_i, \quad (i = 1, 2, \ldots, k).$$

We have then

$$\mathscr{E}(u) = \sum_{i=1}^{k} p_i\left(\sum_{j=1}^{i-1} p_j + \tfrac{1}{2}p_i\right) = \frac{1}{2}.$$

The new variable, u, has thus the same expectation as in the continuous case. The second and higher moments follow from an application of elementary expectation technique. Thus

$$\mathscr{E}(u^r) = \sum_{i=1}^{k} p_i \left(\sum_{j=1}^{i-1} p_j + \tfrac{1}{2} p_i \right)^r \quad (r = 1, 2, 3, \ldots),$$

from which it is easy to deduce that

$$\mathscr{E}(u^2) = \frac{1}{3}\left(1 - \frac{1}{4}\sum_{i=1}^{k} p_i^3\right),$$

$$\mathscr{E}(u^3) = \frac{1}{4}\left(1 - \sum_{i=1}^{k} p_i^3 + \frac{1}{2}\sum_{i=1}^{k} p_i^4 + \sum_{i=1}^{k-1}\sum_{j=i+1}^{k} p_i^3 p_j\right),$$

$$\mathscr{E}(u^4) = \frac{1}{5}\left(1 - \frac{5}{2}\sum_{i=1}^{k} p_i^3 + \frac{5}{2}\sum_{i=1}^{k} p_i^4 - \frac{11}{16}\sum_{i=1}^{k} p_i^5 + 5\sum_{i=1}^{k-1}\sum_{j=i+1}^{k} p_i^3 p_j - \frac{5}{2}\sum_{i=1}^{k-1}\sum_{j=i+1}^{k} p_i^4 p_j \right.$$
$$\left. - \frac{5}{2}\sum_{i=1}^{k-1}\sum_{j=i+1}^{k} p_i^3 p_j^2 - 5\sum_{i=1}^{k-2}\sum_{j=i+1}^{k-1}\sum_{s=j+1}^{k} p_i^3 p_j p_s\right).$$

Correcting to moments about the mean we have

$$\mu_2(u) = \sigma_u^2 = \frac{1}{12}\left(1 - \sum_{i=1}^{k} p_i^3\right),$$

$$\mu_3(u) = \frac{1}{8}\left(-\sum_{i=1}^{k} p_i^3 + \sum_{i=1}^{k} p_i^4 + 2\sum_{i=1}^{k-1}\sum_{j=i+1}^{k} p_i^3 p_j\right),$$

$$\mu_4(u) = \frac{1}{80} + \frac{1}{8}\sum_{i=1}^{k} p_i^3 - \frac{11}{80}\sum_{i=1}^{k} p_i^5 - \frac{1}{2}\sum_{i=1}^{k-1}\sum_{j=i+1}^{k} p_i^4 p_j - \sum_{i=1}^{k-2}\sum_{j=i+1}^{k-1}\sum_{s=j+1}^{k} p_i^3 p_j p_s - \frac{1}{2}\sum_{i=1}^{k-1}\sum_{j=i+1}^{k} p_i^3 p_j^2.$$

It is clear that as k, the number of groups, increases without limit, the moments of u tend to those of a random variable, rectangularly distributed, which is to be expected.

3. The moments of v may be obtained directly in similar fashion, or may be deduced from those of u. We give the first three of these for the sake of completeness, but have omitted the fourth because of its complexity. Thus

$$\mathscr{E}(v) = \frac{1}{2}\left(1 + \sum_{i=1}^{k} p_i^2\right),$$

$$\mathscr{E}(v^2) = \frac{1}{3}\left(1 + 2\sum_{i=1}^{k} p_i^3 + 3\sum_{i=1}^{k-1}\sum_{j=i+1}^{k} p_i p_j^2\right),$$

$$\mathscr{E}(v^3) = \frac{1}{4} - \frac{3}{2}\sum_{i=1}^{k} p_i^2 + 3\sum_{i=1}^{k} p_i^3 - \frac{3}{2}\sum_{i=1}^{k} p_i^4 + \frac{3}{4}\left(\sum_{i=1}^{k} p_i^2\right)^2 + 3\sum_{i=1}^{k-1}\sum_{j=i+1}^{k} p_i p_j^2$$
$$- \sum_{i=1}^{k-1}\sum_{j=i+1}^{k} p_i p_j^3 + 3\sum_{i=1}^{k-2}\sum_{j=i+1}^{k-1}\sum_{s=j+1}^{k} p_i^2 p_j p_s,$$

from which we deduce that

$$\sigma^2(v) = \frac{1}{12} + \frac{2}{3}\sum_{i=1}^{k} p_i^3 - \frac{1}{2}\sum_{i=1}^{k} p_i^2 - \frac{1}{4}\left(\sum_{i=1}^{k} p_i^2\right)^2 + \sum_{i=1}^{k-1}\sum_{j=i+1}^{k} p_i p_j^2,$$

$$\mu_3(v) = -\frac{5}{4}\sum_{i=1}^{k} p_i^2 + 2\sum_{i=1}^{k} p_i^3 - \frac{3}{2}\left(\sum_{i=1}^{k} p_i^4 + \frac{3}{2}\sum_{i=1}^{k} p_i^2\right)^2 + \frac{3}{2}\sum_{i=1}^{k-1}\sum_{j=i+1}^{k} p_i p_j^2 - \sum_{i=1}^{k-1}\sum_{j=i+1}^{k} p_i p_j^3$$
$$+ 3\sum_{i=1}^{k-2}\sum_{j=i+1}^{k-1}\sum_{s=j+1}^{k} p_i^2 p_j p_s + \frac{1}{4}\left(\sum_{i=1}^{k} p_i^2\right)^3 - \left(\sum_{i=1}^{k} p_i^2\right)\left(\sum_{i=1}^{k} p_i^3\right) - \frac{3}{2}\left(\sum_{i=1}^{k} p_i^2\right)\left(\sum_{i=1}^{k-1}\sum_{j=i+1}^{k} p_i p_j^2\right).$$

4. The moments deduced for u and v above are quite general and will hold for any discrete probability distribution. The two discrete distributions in which we are most likely to be interested, however, are the binomial and Poisson's limit to the binomial. We have therefore investigated for these two distributions the departure from rectangularity of the distribution of u as measured by the moment ratios $\sqrt{\beta_1}$ and β_2. The results of these numerical investigations are given in Tables 1 and 2 and are shown graphically in Fig. 1.

Table 1. *$\sqrt{\beta_1}$ and β_2 of the distribution of the transformed variable u when the original variable has distribution function $(q+p)^n$*

p	0·5		0·6		0·7		0·8		0·9	
n	$-\sqrt{\beta_1}$	β_2	$-\sqrt{\beta_1}$	β_2	$-\sqrt{\beta_1}$	β_2	$-\sqrt{\beta_1}$	β_2	$-\sqrt{\beta_1}$	β_2
2	0	2·000	0·029	1·802	0·213	1·468	0·655	1·561	1·590	3·547
4	0	1·831	0·018	1·841	0·044	1·865	0·143	1·655	0·740	1·602
10	0	1·824	0·004	1·824	0·010	1·826	0·060	1·789	0·097	1·802

Table 2. *$\sqrt{\beta_1}$ and β_2 of the distribution of the transformed variable u when the original variable has a Poisson distribution with parameter m*

m	0·2	0·5	0·8	1·0	1·2	1·5	1·7	2·0
$\sqrt{\beta_1}$	1·669	0·586	0·232	0·135	0·087	0·054	0·043	0·033
β_2	3·818	1·581	1·587	1·697	1·769	1·810	1·816	1·817

We may note that as $n \to \infty$ and $p \to 1$ but $n(1-p) = m$, the $\sqrt{\beta_1}$ and β_2 of the binomial distribution tends to the values corresponding to the Poisson distribution with parameter m. If p be fixed then, as $n \to \infty$, each term (except the constants) in the expressions for the moments of u and v will tend separately to zero, and we are left with the moments of the rectangular distribution as might be expected. Fig. 1 shows that the assumption of rectangularity for u may be very misleading, particularly when the asymmetry of the distribution of the original variable is marked. It will be seen from a study of the algebraic expressions for the moments of u that the minimum departure from the rectangle moments will be when all the p_i's are equal, the maximum departure when all but one of the p_i's are zero.

5. R. A. Fisher's application of χ^2 to the probability integral transformation consists in noticing that $-2\log_e u$ is distributed as χ^2 with two degrees of freedom, provided u is a continuous variable. Thus if we have r independent continuous variables $u_1, u_2, \ldots, u_r$, it will follow, owing to the additive property of χ^2, that

$$-2\sum_{i=1}^{r}\log_e u_i$$

is distributed as χ^2 with $2r$ degrees of freedom. The moments of the distribution of a single discontinuous variable, $-\log_e u$, may easily be found. For brevity we write

$$\sum_{j=i+1}^{k} p_j = Y_i,$$

and we have
$$\mathscr{E}(-\log_e u)^b = (-1)^b \sum_{i=1}^{k} p_i[\log_e(1-Y_i-\tfrac{1}{2}p_i)]^b.$$

Using the linear difference operator E we may write

$$\mathscr{E}(\log_e u)^b = \sum_{i=1}^{k} p_i E^{\frac{1}{2}p_i}[\log_e(1-Y_i)]^b,$$

or, since

$$E = 1+\Delta = e^D,$$

$$\mathscr{E}(\log_e u)^b = \sum_{i=1}^{k} p_i e^{\frac{1}{2}p_i D}[\log(1-Y_i)]^b = \sum_{i=1}^{k} p_i \sum_{w=1}^{\infty} \frac{(p_i D)^{w-1}}{2^{w-1}(w-1)!}[\log_e(1-Y_i)]^b.$$

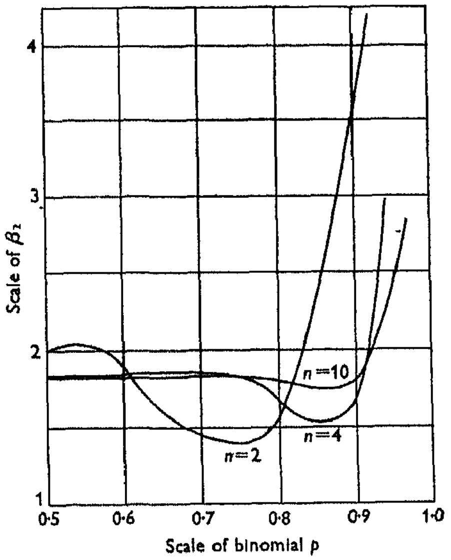

Fig. 1a

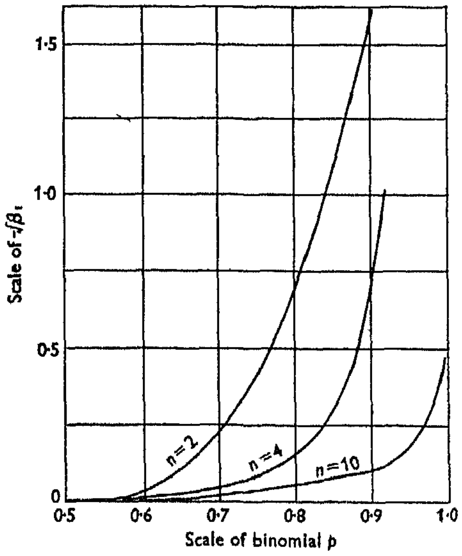

Fig. 1b

This, on substitution and rearrangement, becomes

$$\mathscr{E}(\log_e u)^b = \sum_{w=1}^{\infty} \frac{1}{(w-1)!}\left(\frac{1}{2}\right)^{w-1} \sum_{i=1}^{k} p_i^w \frac{d^{w-1}}{d\left(\sum\limits_{j=i+1}^{k} p_j\right)^{w-1}}\left[\log_e\left(1-\sum_{j=i+1}^{k} p_j\right)\right]^b.$$

This result may be reached alternatively by expanding $\mathscr{E}(\log_e u)$ by the binomial and summing along diagonals. The algebraic expressions for moments do not seem capable of expression in any simpler form. From the expression for $\mathscr{E}(\log_e u)^b$ it is possible to deduce the cumulants of the distribution, and hence, making use of the additive property of cumulants of independent random variables, to write down the cumulants of the distribution of the sum of r independent discontinuous random variables $u_1, u_2, \ldots, u_r$.

6. We now proceed to show that as k, the number of groups, increases without limit, the moments of $-2\log_e u$ tend to those of χ^2 with two degrees of freedom provided each of the p_i are different from zero. We have

$$\lim_{\substack{k\to\infty \\ p_i\to 0 \\ (i=1,2,\ldots,k)}} \mathscr{E}(\log_e u)^b = \lim_{\substack{k\to\infty \\ p_i\to 0 \\ (i=1,2,\ldots,k)}} \sum_{i=1}^{k} p_i\left[\log_e\left(1-\sum_{j=i+1}^{k} p_j - \tfrac{1}{2}p_i\right)\right]^b.$$

Write

$$\sum_{j=1}^{i} p_j = Z_i \quad \text{and} \quad p_i = \Delta Z_i;$$

then

$$\mathscr{E}(\log_e u)^b = \sum_{i=1}^{k} \Delta Z_i[\log(Z_i - \tfrac{1}{2}\Delta Z_i)]^b.$$

It is clear that we may ignore the $\frac{1}{2}\Delta Z_i$ inside the logarithm, since on expansion this will give rise to second-order terms in ΔZ_i which will be negligible in the limit. We take then

$$\lim_{\substack{k\to\infty \\ \Delta Z_i\to 0}} \mathscr{E}(\log_e u)^b = \lim_{\substack{k\to\infty \\ \Delta Z_i\to 0}} \sum_{i=1}^{k} \Delta Z_i[\log_e Z_i]^b = \int_0^1 (\log_e Z)^b\, dZ.$$

Write

$$Z = e^{-z},$$

whence

$$\lim_{\substack{k\to\infty \\ \Delta Z_i\to 0}} \mathscr{E}(\log_e u)^b = (-1)^b \int_0^\infty z^b e^{-z} dz = (-1)^b b!,$$

and therefore

$$\lim_{\substack{k\to\infty \\ \Delta Z_i\to 0}} \mathscr{E}(-2\log_e u)^b = 2^b b!,$$

which is the bth moment about the origin of the distribution of χ^2 with two degrees of freedom.

Table 3. *Actual tail areas of the probability distributions of u and v for the binomial $(q+p)^n$ corresponding to* 0·05 *and* 0·01 *levels in the continuous case*

n	p		0·5	0·6	0·7	0·8	0·9
2	$\chi^2_{0\cdot05}$	u	0	0	0·09	0·04	0·01
		v	0	0	0	0·04	0·01
	$\chi^2_{0\cdot01}$	u	0	0	0	0	0·01
		v	0	0	0	0	0·01
4	$\chi^2_{0\cdot05}$	u	0·0625	0·0256	0·0837	0·0272	0·0523
		v	0	0·0256	0·0081	0·0272	0·0037
	$\chi^2_{0\cdot01}$	u	0	0	0·0081	0·0016	0·0037
		v	0	0	0·0081	0·0016	0·0037
10	$\chi^2_{0\cdot05}$	u	0·0547	0·0548	0·0473	0·0328	0·0702
		v	0·0107	0·0123	0·0473	0·0328	0·0128
	$\chi^2_{0\cdot01}$	u	0·0107	0·0123	0·0106	0·0064	0·0128
		v	0·0010	0·0025	0·0016	0·0064	0·0017

This result is true whatever b. It may be noted that any function of the type

$$\sum_{i=1}^{k} p_i \left[\log_e\left(1 - \sum_{j=i+1}^{k} p_j - Ap_i\right)\right]^b,$$

where A is a constant, will have the same limiting moments.

7. The error involved in assuming that $-2\log_e u$ (or $-2\log_e v$) is distributed as χ^2 with two degrees of freedom was investigated numerically. We have

$$\text{Antilog}_e\left(-\tfrac{1}{2}\chi^2_\epsilon\right) = \epsilon$$

for two degrees of freedom, and we choose ϵ to be 0·05 or 0·01, these being the commonly used significance levels. It is accordingly straightforward to find the actual probability of u (or v) being less than 0·05 or 0·01, the formal significance limits. These probabilities are given in Table 3 for three different values of n and five different values of p for the binomial $(q+p)^n$. It will be seen that when $n = 10$, provided p is not too different from $\frac{1}{2}$, $P\{u \leqslant 0\cdot05\}$ and $P\{u \leqslant 0\cdot01\}$ are not very different from their nominal values.

8. We now turn to the distribution of the discontinuous random variable u, when the mean is fitted from the data. For the Poisson distribution the mean determines all probabilities; for the binomial, $(q+p)^n$, n is often known, but the mean is necessary to determine p and hence the probabilities. We shall therefore discuss the moments of the function

$$u = \sum_{j=1}^{i-1} p_j + \tfrac{1}{2} p_i,$$

when the p's have to be estimated from a random sample of N observations.

It is assumed that the variable may take values $x_1, x_2, \ldots, x_k$. (For example, in the binomial distribution these will be $0, 1, 2, \ldots, k-1$.) The sample consists of N units or observations, n_i of which take the value x_i ($i = 1, 2, \ldots, k$). We let the mean of the sample be X/N or we let

$$n_1 x_1 + n_2 x_2 + \ldots + n_k x_k = X.$$

Define the function $F_r(x_i, X)$ as

$$F_r(x_i, X) = \left(\sum_{j=1}^{i-1} p_j + \tfrac{1}{2} p_i\right)^r.$$

We shall write for brevity

$$F_r(x_i \mid X) = \left(\sum_{j=1}^{i-1} \hat{p}_j + \tfrac{1}{2} \hat{p}_i\right)^r,$$

where $\hat{p}_i$ indicates the value of p_i as estimated from a sample with fixed mean X/N. We have that for all samples

$$\sum_{i=1}^{k} p_i F_r(x_i, X) = \mathscr{E}(F_r(x_i, X)) = \sum_X P\{X\} \mathscr{E}(F_r(x_i \mid X)),$$

the summation being taken over all the possible values of X. Further

$$\mathscr{E}(F_r(x_i \mid X)) = \sum_{i-1}^{k} F_r(x_i \mid X) P\{x_i \mid X\},$$

so that

$$\mathscr{E}(F_r(x_i, X)) = \sum_X \sum_i P\{X\} P\{x_i \mid X\} F_r(x_i \mid X).$$

For a fixed X the probability of getting a partition of N which is $n_1, n_2, \ldots, n_k$, is the multinomial chance

$$\frac{N!}{n_1! n_2! \ldots n_k!} p_1^{n_1} p_2^{n_2} \ldots p_k^{n_k},$$

which we shall write for brevity as $P\{n_1, n_2, \ldots, n_k\}$. It follows that

$$P\{x_i \mid X\} = \frac{P\{x_i\}}{P\{X\}} = \frac{1}{P\{X\}} \sum_X{}^* \frac{n_i}{N} P\{n_1, n_2, \ldots, n_k\},$$

where $\sum_X{}^*$ means the sum over all possible k partitions of N when X is fixed. Substituting this value in $\mathscr{E}(F_r(x_i \mid X))$, we have

$$\mathscr{E}(F_r(x_i \mid X)) = \frac{1}{P\{X\}} \sum_i F_r(x_i \mid X) \sum_X{}^* \frac{n_i}{N} P\{n_1, n_2, \ldots, n_k\}.$$

We now write the multinomial chance as a product,

$$P\{n_1, n_2, \ldots, n_k\} = N! \prod_{s=1}^{k} \frac{p_s^{n_s}}{n_s!}, \tag{α}$$

and we have that

$$\frac{n_i}{N} P\{n_1, n_2, \ldots, n_k\} = (N-1)!\, p_i \prod_{s=1}^{k}{}^{(i)} \frac{p_s^{n_s}}{n_s!}, \tag{β}$$

where the superfix (i) indicates that n_i is to be replaced by $n_i - 1$. Thus

$$\prod_{s=1}^{k}{}^{(i)} \frac{p_s^{n_s}}{n_s!} = \frac{p_1^{n_1}}{n_1!} \frac{p_2^{n_2}}{n_2!} \cdots \frac{p_{i-1}^{n_{i-1}}}{n_{i-1}!} \frac{p_i^{n_i-1}}{(n_i-1)!} \frac{p_{i+1}^{n_{i+1}}}{n_{i+1}!} \cdots \frac{p_k^{n_k}}{n_k!}.$$

It will be remembered that $x_1, x_2, \ldots, x_k$ are fixed values that the variate may take. Since X is fixed we know that the probability of a partition of N into k parts so that

$$\sum_{i=1}^{k} n_i x_i = X \quad \text{and} \quad \sum_{i=1}^{k} n_i = N$$

is given by (α). We have in (β) one variable missing from the ith group. Hence (β) gives us the probability of a partition of N into k parts where

$$\sum_{j=1}^{k}{}^{(i)} n_j x_j = X - x_i \quad \text{and} \quad \sum_{j=1}^{k}{}^{(i)} n_j = N-1.$$

It follows that

$$\sum_X{}^{*} \frac{n_i}{N} P\{n_1, n_2, \ldots, n_k\} = \sum_X{}^{*} (N-1)!\, p_i \prod_{s=1}^{k}{}^{(i)} \frac{p_s^{n_s}}{n_s!}$$
$$= p_i \,.\, P\{\text{sum of } N-1 \text{ variables} = X - x_i\},$$

and that

$$\mathscr{E}(F_r(x_i \mid X)) = \frac{1}{P\{X\}} \sum_{i=1}^{k} F_r(x_i \mid X) \,.\, p_i \,.\, P\{\text{sum of } N-1 \text{ variables} = X - x_i\}.$$

Finally on substitution in $\mathscr{E}(F_r(x_i, X))$, we have

$$\mathscr{E}(F_r(x_i, X)) = \sum_X \sum_{i=1}^{k} F_r(x_i \mid X)\, p_i \,.\, P\{\text{sum of } N-1 \text{ variables} = X - x_i\}$$
$$= \sum_X \sum_{i=1}^{k} \left(\sum_{j=1}^{i-1} \hat{p}_j + \tfrac{1}{2}\hat{p}_i \right)^r . \, p_i \,.\, P\{\text{sum of } N-1 \text{ variables} = X - x_i\}.$$

9. The theory as set out is applicable to any discontinuous variable, the elementary probability law of which is defined by the values $x_1, x_2, \ldots, x_k$, with corresponding probabilities $p_1, p_2, \ldots, p_k$. We now consider the special cases of Poisson and the binomial.

(i) *Poisson's binomial limit*

Assume that a sample of N observations has been drawn from a population which is described by Poisson's binomial limit with unknown parameter m. If the mean of the sample is X/N, then

$$\hat{p}_j = \frac{e^{-X/N}(X/N)^j}{j!},$$

where j may take values $0, 1, 2, \ldots, \infty$. The sum of $N-1$ variables is itself distributed as Poisson with parameter $(N-1)\,m$, and therefore

$$\mathscr{E}(F_r(x_i, X)) = \sum_X \sum_{i=0}^{\infty} \left(\sum_{j=0}^{i-1} \hat{p}_j + \tfrac{1}{2}\hat{p}_i \right)^r \frac{m^i e^{-m}}{i!} \frac{e^{-(N-1)m}\,(m(N-1))^{X-i}}{(X-i)!}.$$

Noticing that

$$(N-1)^{X-i} = N^X \left(1 - \frac{1}{N}\right)^{X-i} \left(\frac{1}{N}\right)^i,$$

after a certain amount of rearrangement we have

$$\mathscr{E}(F_r(x_i, X)) = \sum_X \frac{e^{-Nm}(Nm)^X}{X!} \sum_{i=0}^{\infty} \left(\sum_{j=0}^{i-1} \hat{p}_j + \tfrac{1}{2}\hat{p}_i \right)^r {}^X C_i \left(\frac{1}{N}\right)^i \left(1 - \frac{1}{N}\right)^{X-i},$$

giving the moments of the probability integral variable whose elementary probability law is Poisson with mean estimated from the data.

(ii) *The binomial*

Assume that a sample of N independent observations has been drawn from a binomial population $(q+p)^n$, where the parameter n is known but p is unknown and must be estimated from the data. Here x_j may take values $0, 1, 2, \ldots, n$. The sum of $N-1$ independent binomial variables is itself distributed as a binomial variable with index $(N-1)n$. We have then

$$\hat{p}_j = {}^nC_j\left(\frac{X}{nN}\right)^j\left(1-\frac{X}{nN}\right)^{n-j},$$

and

$$\mathscr{E}(F_r(x_i, X)) = \sum_X\sum_{i=0}^{n}\left(\sum_{j=0}^{i-1}\hat{p}_j+\tfrac{1}{2}\hat{p}_i\right)^r\frac{n!}{i!(n-i)!}p^iq^{n-i}\frac{[(N-1)n]!}{(X-i)!\,((N-1)n-X+i)!}p^{X-i}q^{(N-1)n-X+i},$$

which on rearrangement gives

$$\mathscr{E}(F_r(x_i, X)) = \sum_X {}^{Nn}C_X p^X q^{Nn-X}\sum_{i=0}^{n}\left(\sum_{j=0}^{i-1}\hat{p}_j+\tfrac{1}{2}\hat{p}_i\right)^r\frac{{}^XC_i\,.\,{}^{Nn-X}C_{n-i}}{{}^{Nn}C_n}.$$

Thus, for any discontinuous variable, the moments of the probability integral variable can be worked out when all the parameters of the probability law are known except the mean, which is estimated from the data. From the expressions given it would be possible to decide what the size of the sample must be before the effect of estimating the mean can be neglected, so that the moments approximate closely to those of the distribution of the transformed variable when all the parameters of the probability law of the original variable are known.

REFERENCES

DAVID, F. N. & JOHNSON, N. L. (1948). *Biometrika*, **35**, 182.

LANCASTER, H. O. (1949). *Biometrika*, **36**, 370.

HOW BALANCED INCOMPLETE BLOCK DESIGNS MAY BE MADE TO FURNISH ORTHOGONAL ESTIMATES IN WEIGHING DESIGNS

By K. S. BANERJEE, *Pusa, Bihar, India*

1. Summary

Arrangements given by balanced incomplete block designs may be utilized as efficient spring balance designs (Banerjee, 1948). When these arrangements are used as such as weighing designs, it is not possible to get mutually orthogonal estimates. A method has been detailed here to show how these arrangements may be made to furnish orthogonal estimates.

2. Introduction

Let $x_{i\alpha} = +1, -1$ or 0, if the ith object at the αth weighing is placed in the left pan, right pan or in neither. Here $i = 1, 2, \ldots, p$ and $\alpha = 1, 2, \ldots, N$ $(p \leqslant N)$. Let b_i be the weight of the ith object, y_α the record of the αth weighing when the weight is placed in the right pan (or the negative of the weight placed in the left pan) and c_α a random variable distributed about zero with variance σ^2. The weighing of N linear combinations of p objects in a balance will then result in N linear equations of the form

$$x_{1\alpha}b_1 + x_{2\alpha}b_2 + \ldots + x_{p\alpha}b_p = y_\alpha + c_\alpha. \tag{1}$$

The matrix denoted by $X = [x_{\alpha i}]$ is known as the design matrix. The normal equations giving the least-square estimates may be written in the matrix notation as $[X'X]B = X'Y$, where X' is the transpose of X, B and Y are the column vectors $\{\hat{b}_1 \hat{b}_2 \ldots \hat{b}_p\}$ and $\{y_1 y_2 \ldots y_N\}$ respectively and $\hat{b}_i$ is the least square estimate of b_i. The diagonal elements a^{ii} of the reciprocal matrix $[X'X]^{-1}$ will represent the variance factors, the variance of the ith object being given by $a^{ii}\sigma^2$. The sum of the squares of the residuals divided by the degrees of freedom $(N-p)$ will give an estimate of σ^2.

The estimates are given by $\quad B = [X'X]^{-1}X'Y.$

We have
$$[BB'] = [[X'X]^{-1}X'][[X'X]^{-1}X']' = [X'X]^{-1}X'X[X'X]^{-1} = [X'X]^{-1}.$$

From the above it is clear that if the estimates of the objects are available in mutually orthogonal linear functions of the variables y_α $(\alpha = 1, 2, \ldots, N)$, $[X'X]^{-1}$ will be a diagonal matrix. The constitution of $[X'X]^{-1}$ will therefore reveal if any of the estimates is non-orthogonal to the rest.

In a chemical balance, the object may be placed in both the pans so that the design matrix will have $+1$, -1 and 0 as its elements, but in a spring balance the elements of the design matrix are restricted to be either $+1$ or 0.

Hotelling (1944), the originator of the problem of weighing designs, has shown besides other things that the chemical-balance design of maximum efficiency is obtained when the design matrix X is orthogonal, consisting of ± 1 as its elements. Kishen (1945) pointed out that such a matrix can always be constructed when the number of weighings $N = 2^r$ (r any positive integer). Mood (1946) has pointed out that the construction of the chemical-balance design of maximum efficiency is connected with the Hadamard determinant problem and has shown that the chemical-balance design of maximum efficiency with N weighings exists, if the Hadamard matrix H_N exists. Plackett & Burman (1946) have constructed all Hadamard matrices up to 100 except 92. Mood (1946) has shown further that if a Hadamard matrix H_N exists, the spring-balance design of maximum efficiency also exists. The order of such a design matrix is less than that of H_N by one. Such designs have been denoted as L_N. It has also been proved by Mood that when the number of objects p is odd and $= 2k-1$, the most efficient spring-balance design is obtained when all possible combinations of the objects are weighed k at a time. Such designs have been denoted by Mood as P_k. The present author (1948) has shown that the arrangements afforded by balanced incomplete block designs furnish the same efficiency as the P_k of Mood. He has also pointed out (1948) that the designs L_N are given by a special class of symmetrical balanced incomplete block designs. Kempthorne (1948) doubted whether the optimum spring-balance design L_N of Mood would be practically useful in view of the fact that L_N furnishes correlated estimates. Kempthorne suggested that fractional replicate designs which afford uncorrelated weights were perhaps preferable to the optimum design L_N of Mood. The present author, however, showed (1949) that the design L_N of Mood can also be made to furnish orthogonal estimates, and that these two types of designs are virtually of the same nature. In fact, a slight modification in one gives rise to the other. The nature of the original example of Yates (1935) has also been shown in the note referred to above.

It is now intended to show here that the arrangements given by balanced incomplete block designs in general may also be made to furnish uncorrelated estimates, and that the solution in the case of L_N comes only as a particular instance of the general solution proposed to be discussed here. The identities of Fisher & Yates (1948) well known in the theory of balanced incomplete blocks, $bk = vr$, $\lambda(v-1) = r(k-1)$, have been frequently made use of in the analysis in connexion with this note.

3. WHEN A BALANCED INCOMPLETE BLOCK DESIGN IS USED AS SUCH AS A WEIGHING DESIGN

Usually v and b are used to denote respectively the number of varieties and the number of blocks in balanced incomplete block designs. In weighing designs, v takes the place of p, the number of objects to be weighed, and b that of N, the number of weighings to be made. r and λ, however, have similar meanings to those in connexion with balanced incomplete block designs; that is, r is the number of times each object is weighed and λ the number of times each pair is weighed together. (The analysis henceforward will be written in terms of b and v instead of N and p.) The matrix $[X'X]$ in general will take the form

$$\begin{bmatrix} r & \lambda & \lambda & \dots & \lambda \\ \lambda & r & \lambda & \dots & \lambda \\ \lambda & \lambda & r & \dots & \lambda \\ \dots & \dots & \dots & \dots & \dots \\ \lambda & \lambda & \lambda & \dots & r \end{bmatrix}. \tag{2}$$

(2) is of order $v \times v$. The diagonal elements in the matrix reciprocal to $[X'X]$ will be all positive and equal to

$$\frac{r+\lambda(v-2)}{(r-\lambda)\{r+\lambda(v-1)\}}, \tag{3}$$

while the remaining elements in the reciprocal matrix will be all negative and equal to

$$-\frac{\lambda}{(r-\lambda)\{r+\lambda(v-1)\}}. \tag{4}$$

(4) shows that the estimates will be correlated.

4. WHEN A BALANCED INCOMPLETE BLOCK DESIGN IS ADJUSTED TO SUIT A BIASED SPRING BALANCE

If the balance has a bias, it will be necessary to modify the design matrix to suit the estimation of the bias or that of the weights free of the bias. Taking the bias as an additional object to be weighed, a column of ones and only one row of zeroes in that order may be added to the design matrix to correspond to the bias supposed as an additional object. The design so obtained will then be suitable for the estimation of the weights. This means getting an estimate of the bias by only one weighing operation and subsequently weighing the combinations of the objects along with the bias. If, however a column of ones and t rows of zeroes in that order are added to the design matrix, this will imply devoting t weighings to the estimation of the bias. In such a situation, the matrix $[X'X]$ will be of the form

$$\begin{bmatrix} b+t & r & r & \dots & r \\ r & r & \lambda & \dots & \lambda \\ r & \lambda & r & \dots & \lambda \\ \dots & \dots & \dots & \dots & \dots \\ r & \lambda & \lambda & \dots & r \end{bmatrix}. \tag{5}$$

The order of the above matrix is one more than that of (2) because of the inclusion of the bias and is equal to $(v+1)\times(v+1)$.

I. The value of the determinant

$$\begin{aligned} |X'X| &= (r-\lambda)^{v-1}[(b+t)\{r+\lambda(v-1)\}-r^2v] \\ &= t(r-\lambda)^{v-1}\{r+\lambda(v-1)\}. \end{aligned}$$

II. The value of the determinant obtained after suppressing the first row and the first column of $|X'X|$ is

$$(r-\lambda)^{v-1}\{r+\lambda(v-1)\}.$$

III. The value of the determinant obtained after suppressing the second row and the second column of $|X'X|$ is

$$(r-\lambda)^{v-2}[(b+t)\{r+\lambda(v-2)\}-r^2(v-1)].$$

IV. The value of the determinant obtained after suppressing the first row and the second column of $|X'X|$ is

$$r(r-\lambda)^{v-1}.$$

V. The value of the determinant obtained after suppressing the second row and the third column of $|X'X|$ is

$$(r-\lambda)^{v-2}\{\lambda(b+t)-r^2\}.$$

If the value of t be so chosen that

$$\lambda(b+t)-r^2=0, \quad \text{i.e.} \quad t=\frac{r^2}{\lambda}-b, \tag{6}$$

the value of (V) will be zero and that of (III) will reduce to $t(r-\lambda)^{v-2}\{r+\lambda(v-1)\}$. Using (6) and the identities in the theory of balanced incomplete blocks referred to before, $r-\lambda$ may be shown to be equal to $tk\lambda/r$. The matrix $[X'X]^{-1}$ then reduces to

$$\begin{bmatrix} \frac{1}{t} & -\frac{1}{tk} & -\frac{1}{tk} & \cdots & -\frac{1}{tk} \\ -\frac{1}{tk} & \frac{r}{tk\lambda} & 0 & \cdots & 0 \\ -\frac{1}{tk} & 0 & \frac{r}{tk\lambda} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ -\frac{1}{tk} & 0 & 0 & \cdots & \frac{r}{tk\lambda} \end{bmatrix} \tag{7}$$

(7) shows that except for the estimate of the bias, the other estimates will be mutually orthogonal. The estimates of the bias as well as those of the objects will be available as

$$[X'X]^{-1}X' = \begin{bmatrix} \frac{1}{t} & \frac{1}{t} & \frac{1}{t} & 0 & 0 & 0 & \cdots \\ -\frac{1}{tk} & -\frac{1}{tk} & -\frac{1}{tk} & \frac{1}{r} & \frac{1}{r} & \frac{1}{r} & \cdots \\ -\frac{1}{tk} & -\frac{1}{tk} & -\frac{1}{tk} & \frac{1}{r} & -\frac{1}{tk} & -\frac{1}{tk} & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix} \tag{8}$$

For every 0 and 1 in all rows of X' except in the first, there will respectively be $-1/tk$ and $1/r$ in all the rows of (8) except in the first. The elements in the rows of (8) are the coefficients of y_α in the linear functions estimating the weights of the objects. It is easy to see that the sum product of the coefficients in any two linear functions estimating two objects is equal to

$$\frac{t}{t^2k^2}+\lambda\left(\frac{1}{r^2}\right)+\lambda_0\left(\frac{1}{t^2k^2}\right)+(b-\lambda-\lambda_0)\left(\frac{1}{r}\right)\left(-\frac{1}{tk}\right), \tag{9}$$

where $\lambda_0 = b-2r+\lambda$. (9) on simplification may be shown to be equal to zero.

5. SITUATION IN A COMPLEMENTARY BALANCED INCOMPLETE BLOCK DESIGN

Given a balanced incomplete block design with parameters v, b, r, k, λ, it is always possible to obtain another with parameters

$$v_0 = v, \quad b_0 = b, \quad r_0 = b-r, \quad k_0 = v-k, \quad \lambda_0 = b-2r+\lambda,$$

by replacing each block by its complement. Two such designs are called complementary to one another. The distributions of ones and of zeroes in X furnish the two designs, which are complementary to one another.

When an integral value of t is available as given by (6), it is possible to get the orthogonal estimates also from the complementary design. In this situation a column of ones and t rows of ones in that order have to be added to the design matrix X to achieve the desired object. The matrix $[X'X]$ will then take the same form as (5), where $r' = r_0+t$ and $\lambda' = \lambda_0+t$ are substituted respectively for r and λ. The value of the determinant obtained by suppressing the second row and the third column of $|X'X|$ in this case will be given by

$$(r'-\lambda')^{v-2}\{\lambda'(b+t)-r'^2\}. \tag{10}$$

It is easy to see that (10) vanishes, as

$$\begin{aligned}\lambda'(b+t)-r'^2 &= (\lambda_0+t)(b+t)-(r_0+t)^2 \\ &= (b-2r+\lambda+t)(b+t)-(b-r+t)^2 \\ &= 0, \quad \text{by (6)}.\end{aligned}$$

It can easily be shown by using (6) and the parametric relations that

$$r' = tk, \quad \lambda' = \frac{\lambda(tk)^2}{r^2}, \quad r'-\lambda' = \frac{tk\lambda}{r} \quad \text{and} \quad r'+\lambda'(v-1) = tk\left\{1+\frac{tk(k-1)}{r}\right\}.$$

The matrix $[X'X]^{-1}$ in this situation will then reduce to the form

$$\begin{bmatrix} \phi & -\frac{1}{r} & -\frac{1}{r} & \cdots & -\frac{1}{r} \\ -\frac{1}{r} & \frac{r}{tk\lambda} & 0 & \cdots & 0 \\ -\frac{1}{r} & 0 & \frac{r}{tk\lambda} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ -\frac{1}{r} & 0 & 0 & \cdots & \frac{r}{tk\lambda} \end{bmatrix},$$

where $\phi = \frac{1}{r}\left\{1+\frac{tk(k-1)}{r}\right\}$. The estimates of the objects will be given by

$$[X'X]^{-1}X' = \begin{bmatrix} \phi-\frac{v}{r} & \phi-\frac{v}{r} & \phi-\frac{v}{r} & \phi-\frac{k_0}{r} & \phi-\frac{k_0}{r} & \phi-\frac{k_0}{r} & \cdots \\ \frac{1}{tk} & \frac{1}{tk} & \frac{1}{tk} & -\frac{1}{r} & -\frac{1}{r} & -\frac{1}{r} & \cdots \\ \frac{1}{tk} & \frac{1}{tk} & \frac{1}{tk} & -\frac{1}{r} & \frac{1}{tk} & \frac{1}{tk} & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix} \tag{11}$$

For every 0 and 1 in all the rows of X' except in the first, there will respectively be $-1/r$ and $1/tk$ in all the rows of (11) except in the first. The sum product of the coefficients in any two linear functions estimating two objects will be exactly the same as (9).

It may be noticed that the variance factor for an object in the complementary design will be the same as that of the original design.

In the case of L_N as given by a symmetrical balanced incomplete block design, $t = 1$, so that only one additional row has to be added to the design matrix to suit the estimation of the bias. The treatment given in the note referred to before (Banerjee, 1949) in respect to L_N is, therefore, obtained as a special case of the general procedure outlined here.

6. SOLUTION OF A MORE GENERAL NATURE

The above procedure will fail to furnish orthogonal estimates when r^2/λ is not an integral number. A solution by the above procedure, however, always exists for all designs when $\lambda = 1$. When r^2 is not divisible by λ, we can find the least integer ξ such that $(r+\xi)^2$ is divisible by $\lambda+\xi$. If a column of ones and ξ rows of ones and η rows of zeroes in that order are added on to the design matrix to suit the estimation of the bias in a biased spring balance, the matrix $[X'X]$ will be of the same form as (5), where the left top element will be $b+\xi+\eta$, and r and λ

will respectively be substituted by $r'' = r+\xi$ and $\lambda'' = \lambda+\xi$. The value of the determinant $|X'X|$ obtained by suppressing the second row and the third column will be equal to

$$(r''-\lambda'')^{v-2}\{(b+\xi+\eta)(\lambda+\xi)-(r+\xi)^2\}. \tag{12}$$

Now (12) = 0 if

$$(b+\xi+\eta)(\lambda+\xi)-(r+\xi)^2 = 0. \tag{13}$$

Knowing that $(r+\xi)^2$ is a multiple of $(\lambda+\xi)$, the value of η will be determined by the condition (13).

The results corresponding to I, II, III and IV of § 4 reduce respectively to

$$(r-\lambda)^v(b+\xi+\eta),\quad (r-\lambda)^{v-1}\{r+\lambda(v-1)+\xi v\},$$
$$(r-\lambda)^{v-1}(b+\xi+\eta)\quad\text{and}\quad (r+\xi)(r-\lambda)^{v-1}.$$

The matrix $[X'X]^{-1}$ will therefore be obtained as

$$\begin{bmatrix} A & -C & -C & \ldots & -C \\ -C & \dfrac{1}{r-\lambda} & 0 & \ldots & 0 \\ -C & 0 & \dfrac{1}{r-\lambda} & \ldots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ -C & 0 & 0 & \ldots & \dfrac{1}{r-\lambda} \end{bmatrix},$$

where
$$A = \frac{r+\lambda(v-1)+\xi v}{(b+\xi+\eta)(r-\lambda)} \quad\text{and}\quad C = \frac{r+\xi}{(b+\xi+\eta)(r-\lambda)}.$$

The above matrix is of order $(v+1)\times(v+1)$. The estimates of the bias as also those of the objects will be given by

$$[X'X]^{-1}X'Y = \begin{bmatrix} \eta\text{ terms} & \xi\text{ terms} & & & & \\ D\ldots D & A\ldots A & E & E & E & \ldots \\ F\ldots F & -C\ldots -C & F & F & F & \ldots \\ F\ldots F & -C\ldots -C & F & F & -C & \ldots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix} \begin{bmatrix} \text{The column vector} \\ \text{of the records of} \\ \text{weighings, which} \\ \text{are } (b+\xi+\eta) \text{ in} \\ \text{number} \end{bmatrix}, \tag{14}$$

where
$$D = A-vC,\quad E = A-kC \quad\text{and}\quad F = -C+\frac{1}{r-\lambda}.$$

For every 0 and 1 in all the rows of X' except the first, there will respectively be F and $-C$ in all the rows of $[X'X]^{-1}X'$ of (14) except the first. The elements in the rows of the matrix $[X'X]^{-1}X'$ are the coefficients of the records of weighings in the linear expressions estimating the weights of the objects.

The above analysis shows that the more general class of balanced incomplete block designs may also be made to furnish orthogonal estimates. It is also possible to get orthogonal estimates from a design, which is complementary to this design. To do this we add to the design matrix η rows of ones and ξ rows of zeroes instead of adding ξ rows of ones and η rows of zeroes as was done in the former situation. The proof follows from the fact that

$$\begin{aligned}(b+\eta+\xi)(\lambda_0+\eta)-(r_0+\eta)^2 &= (b+\eta+\xi)(b-2r+\lambda+\eta)-(b-r+\eta)^2 \\ &= 0, \quad\text{by (13)}.\end{aligned}$$

7. NUMERICAL ILLUSTRATIONS

The design matrix X in an L_7 (i.e. in a symmetrical balanced incomplete block design afforded by $v = 7, b = 7, r = 4, k = 4, \lambda = 2$) is given by

$$X = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

where the rows refer to the weighing operations and the columns to the objects $b_1, b_2, \ldots, b_7$ to be weighed.

Seven small copper pieces have been arbitrarily chosen as the objects for the purpose of this illustration. The results of weighings (in grams) of the combinations of objects as given in the above design matrix are 10·76251, 7·83798, 6·11380, 12·07808, 8·90452, 10·63856 and 12·80768. (The fifth decimal place of these records has been determined by the method of oscillation, and no temperature and vacuum corrections have been made as the purpose was not to find the accurate weights of any specific objects.)

The left-hand side of the normal equations will be given by a square matrix $[X'X]$ of order 7×7, whose diagonal elements are 4 and the elements elsewhere are 2. The right-hand side (i.e. the numerical side) of the normal equations is given by

$$Z_1 = 46{\cdot}28683, \quad Z_2 = 41{\cdot}62826, \quad Z_3 = 38{\cdot}14357, \quad Z_4 = 38{\cdot}46456,$$
$$Z_5 = 37{\cdot}85891, \quad Z_6 = 36{\cdot}66842 \quad \text{and} \quad Z_7 = 37{\cdot}52197.$$

Subtracting from each of these one-eighth of their total, and dividing by 2, the estimates are obtained as

$$\hat{b}_1 = 5{\cdot}85763, \quad \hat{b}_2 = 3{\cdot}52835, \quad \hat{b}_3 = 1{\cdot}78600, \quad \hat{b}_4 = 1{\cdot}94650,$$
$$\hat{b}_5 = 1{\cdot}64367, \quad \hat{b}_6 = 1{\cdot}04843 \quad \text{and} \quad \hat{b}_7 = 1{\cdot}47520.$$

The matrix $[X'X]$ shows that the estimates are correlated. They may, however, be obtained as orthogonal functions of the records of weighings, if the weighings are made in a biased spring balance. It has been pointed out in § 5 that in the case of L_N, $t = 1$, so that a column of ones and only one row of zeroes have to be added (in that order) to the design matrix in order to get orthogonal estimates. This means that the measure of the bias of the spring balance has to be obtained by only one weighing operation and the remaining combinations have to be weighed as such in the biased balance. The records of weighings and the modified design matrix are shown below, the bias being denoted by b_0.

$$\begin{array}{cccccccccc} b_0 & b_1 & b_2 & b_3 & b_4 & b_5 & b_6 & b_7 & & \text{g.} \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & = & 0{\cdot}00101 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & = & 10{\cdot}76379 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & = & 7{\cdot}83990 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & = & 6{\cdot}11580 \\ 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & = & 12{\cdot}08000 \\ 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & = & 8{\cdot}90543 \\ 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & = & 10{\cdot}63998 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & = & 12{\cdot}80998 \end{array} \qquad (15)$$

The solution of the normal equations giving the estimates is straightforward. The constitution of the matrix $[X'X]^{-1}X'Y$ assists in appreciating how the estimates are obtained as orthogonal linear functions. On substitution of the values of the parameters, this matrix reduces to

$$\begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ -\frac{1}{4} & +\frac{1}{4} & -\frac{1}{4} & \dots & +\frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & +\frac{1}{4} & \dots & +\frac{1}{4} \\ \dots & \dots & \dots & \dots & \dots \\ -\frac{1}{4} & +\frac{1}{4} & +\frac{1}{4} & \dots & +\frac{1}{4} \end{bmatrix} \begin{bmatrix} \text{The column vector} \\ \text{of the 8 records} \\ \text{of weighings in} \\ \text{(15)} \\ \end{bmatrix} \qquad (16)$$

The above furnishes the estimates as

$$\hat{b}_0 = 0{\cdot}00101, \quad \hat{b}_1 = 5{\cdot}85790, \quad \hat{b}_2 = 3{\cdot}52868, \quad \hat{b}_3 = 1{\cdot}78558,$$
$$\hat{b}_4 = 1{\cdot}94662, \quad \hat{b}_5 = 1{\cdot}64354, \quad \hat{b}_6 = 1{\cdot}04887 \quad \text{and} \quad \hat{b}_7 = 1{\cdot}47576.$$

The calculation leading to the solution may be simplified as follows:

Let -1 be put for every 0 in all the columns of the matrix (15) except in the first. The columns then consist only of $+1$ and -1 as their elements. The algebraic sum of the records of weights corresponding to the minus and plus signs of a particular column divided by 4 (i.e. k) will give the estimate of that object which the column represents.

The balanced incomplete block design given by

$$r = 7, \quad v = 8, \quad k = 4, \quad b = 14, \quad \lambda = 3,$$

will not furnish orthogonal estimates when used as such as a spring-balance design. Here, also, orthogonal estimates may be obtained, if the design matrix is modified to suit the estimation of the weights in a biased spring balance as suggested in § 6. In the present instance, $\xi = 1$, $\eta = 1$. Records of weights and the modified design matrix are shown below. An additional copper piece was taken to make the number of objects equal to 8.

$$\begin{array}{ccccccccccc} b_0 & b_1 & b_2 & b_3 & b_4 & b_5 & b_6 & b_7 & b_8 & & \text{g.} \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & = & 18{\cdot}91670 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & = & 0{\cdot}00101 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & = & 13{\cdot}12036 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & = & 12{\cdot}49119 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & = & 10{\cdot}32320 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & = & 10{\cdot}32997 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & = & 5{\cdot}79780 \\ 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & = & 6{\cdot}42631 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & = & 8{\cdot}59500 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & = & 8{\cdot}58860 \\ 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & = & 12{\cdot}07940 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & = & 10{\cdot}76320 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & = & 11{\cdot}08093 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & = & 6{\cdot}83861 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & = & 8{\cdot}15469 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & = & 7{\cdot}83968 \end{array} \qquad (17)$$

When the numerical values of the parameters are substituted, the matrix $[X'X]^{-1}X'Y$ takes the following form:

$$\begin{bmatrix} -\frac{7}{16} & \frac{9}{16} & \frac{1}{16} & \frac{1}{16} & \cdots & \frac{1}{16} \\ +\frac{1}{8} & -\frac{1}{8} & +\frac{1}{8} & +\frac{1}{8} & \cdots & -\frac{1}{8} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ +\frac{1}{8} & -\frac{1}{8} & -\frac{1}{8} & +\frac{1}{8} & \cdots & -\frac{1}{8} \end{bmatrix} \begin{bmatrix} \text{The column vector} \\ \text{of the 16 records} \\ \text{of weighings in} \\ \text{(17)} \end{bmatrix}$$

The estimates are obtained as

$$\hat{b}_0 = 0{\cdot}00132, \quad \hat{b}_1 = 5{\cdot}85791, \quad \hat{b}_2 = 3{\cdot}52807, \quad \hat{b}_3 = 1{\cdot}78583, \quad \hat{b}_4 = 1{\cdot}94731,$$
$$\hat{b}_5 = 1{\cdot}64365, \quad \hat{b}_6 = 1{\cdot}04861, \quad \hat{b}_7 = 1{\cdot}47471 \quad \text{and} \quad \hat{b}_8 = 1{\cdot}62960.$$

Here, also, the calculations may be simplified as in the above instance. In this case, the algebraic sum has to be divided by 8 instead of 4 to get the estimates.

No attempt has been made here to estimate the error from the observations. There will be no degrees of freedom left for the estimation of the error variance in case of L_7, as only seven observations were utilized in estimating the weights of seven objects. Only 1 and 7 D.F. will be available for the estimation of the variances in the other two illustrations.

The numerical illustrations have been added subsequently to this note according to the suggestion of a referee, to whom my grateful thanks are due.

REFERENCES

BANERJEE, K. S. (1948). Weighing designs and balanced incomplete blocks. *Ann. Math. Statist.* **19**, 394.

BANERJEE, K. S. (1949). A note on weighing designs. *Ann. Math. Statist.* **20**, 300.

FISHER, R. A. & YATES, F. (1948). *Statistical Tables for Biological, Agricultural and Medical Research*, Third Ed., p. 17. London: Oliver and Boyd.

HOTELLING, HAROLD (1944). Some improvements in weighing and other experimental techniques. *Ann. Math. Statist.* **15**, 297.

KEMPTHORNE, O. (1948). The factorial approach to the weighing problem. *Ann. Math. Statist.* **19**, 238.

KISHEN, K. (1945). On the design of experiments of weighing. *Ann. Math. Statist.* **16**, 294.

MOOD, A. M. (1946). On Hotelling's weighing problem. *Ann. Math. Statist.* **17**, 432.

PLACKETT, R. L. & BURMAN, J. P. (1946). The design of optimum multifactorial experiments. *Biometrika*, **33**, 305.

YATES, F. (1935). Complex experiments. *J.R. Statist. Soc. Suppl.* **2**, 181 (211).

THE POWER OF THE χ^2 INDEX OF DISPERSION TEST WHEN NEYMAN'S CONTAGIOUS DISTRIBUTION IS THE ALTERNATE HYPOTHESIS

By G. I. BATEMAN

1. The index of dispersion is commonly used to test for the equality of variance and mean under the hypothesis that observations have come from a Poisson population. Thus if there are N random independent variables $x_1, x_2, \ldots, x_N$, each following the same Poisson law, it is known that the quantity

$$z = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{\bar{x}} \tag{1}$$

is approximately distributed as χ^2 with $N-1$ degrees of freedom. The exact distribution of z is not known. Hoel (1943) found the first four moments of the distribution, using Fisher's k-statistic technique. If m is the population mean of the Poisson distribution, he showed that the moments of z are

$$\left.\begin{aligned}
\mu_1'(z) &= N-1, \\
\mu_2(z) &= 2(N-1)\left[1 - \frac{1}{Nm} - \frac{1}{(Nm)^2} - \ldots\right], \\
\mu_3(z) &= 8(N-1)\left[1 + \frac{1}{2m} - \frac{4}{Nm} + \ldots\right], \\
\mu_4(z) &= 12(N-1)\left[N + 3 + \frac{6}{m} + \frac{2}{3m^2} + \ldots\right],
\end{aligned}\right\} \tag{2}$$

where the terms in square brackets are correct to $1/(Nm)^2$, $1/Nm$, and $1/(Nm)^0$ respectively. Hoel carried out a numerical investigation and reached the conclusion that for $m \geqslant 5$ the χ^2 approximation to the distribution of z is highly satisfactory, even for N, the number of observations as small as 5, and is fairly accurate for $m < 5$, although it will tend to give too few significant results.

2. Suppose now that the probability law of the observations, x, is represented by Neyman's contagious distribution. This may be written as

$$\left.\begin{aligned}
P\{x_i = 0\} &= \exp\left(-m_1(1-e^{-m_2})\right) \\
P\{x_i = t+1\} &= \frac{m_1 m_2 e^{-m_2}}{t+1} \sum_{j=0}^{t} \frac{m_2^j}{j!} P\{x_i = t-j\}
\end{aligned}\right\} \quad (\text{for } i = 1, 2, \ldots, N;\ t = 0, 1, 2, \ldots), \tag{3}$$

where

$$m_1 m_2 = \mu_1',$$

$$m_1 m_2 (1 + m_2) = \mu_2.$$

When there is no 'contagion' present in the material observed, then

$$m_1 \to \infty, \quad m_2 \to 0 \quad \text{and} \quad m_1 m_2 = m,$$

the distribution becoming that of Poisson with parameter, m. The moments of the index of dispersion, z, of (1), under the hypothesis described by (3) may be found using the same technique as did Hoel. Thus we have

$$\mu_1'(z) = (N-1)\left[1+m_2-\frac{m_2}{Nm_1m_2}-\frac{m_2^2}{(Nm_1m_2)^2}-\frac{2m_2^3}{(Nm_1m_2)^3}+\dots\right],$$

$$\mu_2(z) = (N-1)\left[2(1+m_2)^2+\frac{m_2(2+m_2)}{m_1m_2}-\frac{1}{Nm_1m_2}\left(2+16m_2+15m_2^2+2m_2^3+\frac{m_2(2-m_2^2)}{m_1m_2}\right)\right.$$
$$\left.-\frac{1}{(Nm_1m_2)^2}\left(2-10m_2-20m_2^2+5m_2^3+2m_2^4+\frac{m_2(2+4m_2+m_2^2-2m_2^3)}{m_1m_2}\right)+\dots\right],$$

$$\mu_3(z) = (N-1)\left[8(1+m_2)^3+\frac{4}{m_1m_2}\left(1+12m_2+20m_2^2+9m_2^3+m_2^4+\frac{m_2(4+8m_2+m_2^2)}{4m_1m_2}\right)\right.$$
$$-\frac{4}{Nm_1m_2}\left(8+60m_2+103m_2^2+57m_2^3+8m_2^4+\frac{m_2}{2m_1m_2}(79+248m_2+160m_2^2+24m_2^3)\right)$$
$$\left.+\frac{m_2^2(9-5m_2)}{2(m_1m_2)^2}+\dots\right],$$

$$\mu_4(z) = 12(N-1)^2\left[(1+m_2)^2+\frac{m_2(2+m_2)}{2m_1m_2}\right]^2$$
$$+12(N-1)\left[4(1+m_2)^4+\frac{1}{m_1m_2}(6+60m_2+147m_2^2+136m_2^3+49m_2^4+6m_2^5)\right.$$
$$+\frac{1}{6(m_1m_2)^2}(4+168m_2+694m_2^2+828m_2^3+371m_2^4+72m_2^5+4m_2^6)$$
$$\left.+\frac{m_2}{12(m_1m_2)^3}(8+56m_2+48m_2^2+25m_2^3+9m_2^4)\right]+\dots. \tag{4}$$

If we put $m_2 = 0$ and $m_1m_2 = m$ these moments reduce to those given in (2).

3. These expressions for the moments of z are complicated in form, but there is an immediate simplification if either m_1m_2, or N, or both, is sufficiently large. The first four moments, as well as being correct to terms of order N^{-2}, N^{-1}, N^0 and N, respectively, are also correct to terms of order $(m_1m_2)^{-3}$, $(m_1m_2)^{-2}$, $(m_1m_2)^{-1}$ and $(m_1m_2)^0$, respectively. Hence, for m_1m_2 sufficiently large, we shall have

$$\left.\begin{aligned}&\mu_1'(z) \doteq (N-1)(1+m_2), \quad \mu_2(z) \doteq 2(N-1)(1+m_2)^2,\\ &\mu_3(z) \doteq 8(N-1)(1+m_2)^3, \quad \mu_4(z) \doteq 12(N-1)^2(1+m_2)^4+48(N-1)(1+m_2)^4.\end{aligned}\right\} \tag{5}$$

These are the moments of $(1+m_2)\chi^2$, where χ^2 has $N-1$ degrees of freedom. We may assume therefore that for Neyman's contagious distribution the index of dispersion is distributed as $\chi^2(1+m_2)$, provided the population mean is sufficiently large. It often happens in bacteriological work that this is the case (see, for example, R. A. Fisher, *Statistical Methods for Research Workers*, § 16). In field ecology, when counts are made on an abundant species, and the quadrat used is not too small, this condition is again realized.

4. If N is large then, regardless of the size of m_1m_2, we have, from the first two equations of (4),

$$\mu_1'(z) \doteq (N-1)(1+m_2), \quad \mu_2(z) \doteq 2(N-1)(1+m_2)^2\left(1+\frac{m_2(2+m_2)}{2m_1m_2(1+m_2)^2}\right), \tag{6}$$

and $\beta_1(z)$ and $\beta_2(z)$ tend to 0 and 3 respectively as N increases without limit. Thus we may assume that

$$\frac{z-(N-1)(1+m_2)}{\left(2(N-1)(1+m_2)^2\left(1+\frac{m_2(2+m_2)}{2m_1m_2(1+m_2)^2}\right)\right)^{\frac{1}{2}}} \tag{7}$$

tends to be normally distributed with zero mean and unit variance as $N \to \infty$. If, in addition, $m_1 m_2$ is large,

$$\frac{z-(N-1)(1+m_2)}{(1+m_2)(2(N-1))^{\frac{1}{2}}} \tag{8}$$

tends to be normally distributed with zero mean and unit variance. This last is of course another way of saying that, as $N \to \infty$, $(1+m_2)\chi^2$, where χ^2 has $N-1$ degrees of freedom, tends to be normally distributed. A closer approximation will be to make use of the fact that, for large N, $\sqrt{(2\chi^2)}$ tends to be normally distributed with mean $\sqrt{[2(N-1)-1]}$ and unit variance, and we have in fact used this approximation below to obtain

$$P\{(1+m_2)\chi^2 > \chi^2_{0\cdot05}\},$$

when $N > 100$ and $\chi^2_{0\cdot05}$ is the value of χ^2 with $N-1$ degrees of freedom cutting off 0·05 from the right-hand tail area. We note further that (7) differs very little from (8) even when $m_1 m_2$ is not large, the difference in denominators being less than 5 % for $m_1 m_2 = 5$. Thus if N is large, even if $m_1 m_2$ is not large, the distribution of z should be well approximated by that of $(1+m_2)\chi^2$.

5. Calculations were carried out with a view to determining how large $m_1 m_2$ should be, when N is not necessarily large, and m_2 is given, for the χ^2 approximation based on the moments (5) to hold. The ratios of the first four moments of z to the corresponding moments of $(1+m_2)\chi^2$ are referred to as R_1, R_2, R_3, R_4, respectively, and the values of these ratios for certain values of $m_1 m_2$, m_2 and N are given in Table 1. From a study of this table we notice that in no case do the first two moments differ greatly from those of $(1+m_2)\chi^2$.* The most extreme case shown is that for which $m_1 m_2 = 5$, $m_2 = 2$, and $N = 5$ and in this case the means differ by 3 % and the standard deviations by 8 %. In studying the ratios R_3 and R_4 it should be borne in mind that $\beta_1(\chi^2) \to 0$ and $\beta_2(\chi^2) \to 3$ as $N \to \infty$. Now

$$\beta_1(z) = \frac{R_3^2}{R_2^3}\beta_1(\chi^2) \quad \text{and} \quad \beta_2(z) = \frac{R_4}{R_2^2}\beta_2(\chi^2),$$

and it follows therefore that $\beta_1(z) \to 0$ and $\beta_2(z) \to 3$ as $N \to \infty$, since $R_4/R_2^2 \to 1$. Thus although R_3 increases with N, yet the $(\beta_1 \beta_2)$ point for z, if N is large, will be close to that of χ^2 with $N-1$ degrees of freedom as we have pointed out previously. We regard the position as satisfactory for $N = 50$ and $m_1 m_2 = 5$ and $m_2 = 1$. For the same N and larger $m_1 m_2$ or smaller m_2 the position will be even better.

6. As pointed out above, Neyman's contagious distribution of (3) may be regarded as depending on two parameters $m_1 m_2 = m$ and m_2, where $1+m_2$ is the ratio of the variance to the mean. If $m_2 = 0$ and m is fixed it becomes the Poisson distribution. Suppose now it is wished to test the hypothesis that $m_2 = 0$ against an admissible set of alternative hypotheses defined by (3), using a significance level α. We may do this by using the index of dispersion, z. The power of this test with regard to a specific alternative defined by $m_2 = M_2$ will then be, approximately,

$$P\{z \geqslant \chi^2_\alpha \mid m_2 = M_2\} = P\{(1+M_2)\chi^2 \geqslant \chi^2_\alpha\}, \tag{9}$$

where χ^2 follows the standard distribution with $N-1$ degrees of freedom and χ^2_α is the significance level α of standard tables. Fig. 1 shows the power obtained in this way, plotted against $(1+m_2)$ for a number of values of N. The curves must be read bearing in mind the restrictions imposed in approximating to the distribution of $(1+m_2)\chi^2$. For $N = 50$, and upwards, it

---

* It is doubtful whether the order of approximation used in obtaining the moments of z is sufficient to give R_3 and R_4 to first decimal place accuracy for cases where $(m_1 m_2)$ is small and m_2 large. These values should therefore be treated with reserve. They have been included because they serve to show the general trend.

Table 1. *The ratios R_1, R_2, R_3, R_4 of the first four moments of the Poisson index of dispersion, z, to the corresponding moments of $(1+m_2)\chi^2$, where χ^2 has $N-1$ degrees of freedom. The probability law for the observations is Neyman's contagious distribution with $\mu_1' = m_1m_2$, $\mu_2' = m_1m_2(1+m_2)$*

m_2	N	$m_1m_2=5$				$m_1m_2=10$				$m_1m_2=20$			
		R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4
0	5	1	0·958	0·96	1·08	1	0·980	0·97	1·04	1	0·980	0·99	1·02
	10	1	0·980	1·02	1·07	1	0·990	1·01	1·04	1	0·995	1·01	1·02
	20	1	0·990	1·06	1·05	1	0·995	1·03	1·02	1	0·998	1·02	1·01
	∞	1	1	1·1	1	1	1	1·05	1	1	1	1·025	1
0·5	5	0·9864	0·932	0·92	1·40	0·9933	0·986	0·97	1·18	0·9966	0·983	0·99	1·09
	10	0·9933	0·993	1·15	1·35	0·9966	0·997	1·08	1·16	0·9983	0·998	1·04	1·08
	20	0·9966	1·024	1·27	1·26	0·9983	1·012	1·14	1·12	0·9991	1·006	1·07	1·06
	50	0·9987	1·043	1·35	1·18	0·9993	1·022	1·17	1·09	0·9997	1·011	1·09	1·04
	∞	1	1·0556	1·3969	1·1142	1	1·0278	1·1969	1·0563	1	1·0139	1·0981	1·0280
1·0	5	0·9792	0·903	0·87	1·57	0·9898	0·951	0·95	1·26	0·9949	0·975	0·98	1·12
	10	0·9898	0·988	1·20	1·49	0·9949	0·994	1·11	1·23	0·9975	0·997	1·06	1·11
	20	0·9949	1·031	1·36	1·37	0·9975	1·016	1·19	1·17	0·9987	1·008	1·10	1·08
	50	0·9980	1·057	1·47	1·25	0·9990	1·029	1·24	1·12	0·9995	1·014	1·12	1·06
	100	0·9990	1·066	1·51	1·21	0·9995	1·033	1·25	1·10	0·9998	1·017	1·13	1·05
	∞	1	1·0750	1·5456	1·1556	1	1·0375	1·2708	1·0764	1	1·0188	1·1349	1·037
2·0	5	0·9709	0·849	0·75	1·79	0·9861	0·923	0·91	1·35	0·9932	0·961	0·96	1·16
	10	0·9861	0·968	1·22	1·66	0·9932	0·984	1·13	1·30	0·9966	0·992	1·07	1·14
	20	0·9932	0·972	1·47	1·49	0·9966	1·014	1·24	1·23	0·9983	1·007	1·12	1·11
	50	0·9971	1·065	1·62	1·32	0·9987	1·032	1·31	1·15	0·9993	1·016	1·16	1·07
	100	0·9987	1·077	1·67	1·26	0·9993	1·038	1·34	1·12	0·9997	1·022	1·17	1·06
	∞	1	1·0889	1·7237	1·1857	1	1·0444	1·3596	1·0900	1	1·0222	1·1793	1·044

will be seen that we are concerned with only small values of m_2. It might therefore safely be assumed from the evidence of Table 1 that these power curves will be a good representation of the power of z when the mean is as small as 5, and for even smaller means as N increases. For the curves for small values of N, i.e. 5, 10, 20, the mean m_1m_2 must be large for the $(1+m_2)\chi^2$ approximation to hold, unless m_2 is small. For high values of m_2, i.e. when the power is large, it would be necessary (vide Table 1) for $m_1m_2 > 20$ for this part of the curves to be accurate. Apart from these restrictions, we may note that the power is independent of the mean of the population.

7. The $(1+m_2)\chi^2$ approximation to the distribution of z is simple to handle and can readily be used, before the collection of data, to assess the size of sample required in order to have a given chance of detecting a departure from randomness due to clustering of Neyman's 'contagious' kind. Thus, using the 5 % significance level, a sample of ten observations from a Neyman's contagious population will only enable departures from the Poisson to be detected nine times out of ten when the variance of the distribution is actually four times the mean. In order to detect a 2 : 1 ratio with the same probability it is necessary to have

a sample of between thirty and forty. Large numbers of observations, but not large means, are comparatively rare in cases where the index of dispersion is used, and it is probably for this reason that clustering effects are very often overlooked and the hypothesis of randomness, as defined by the Poisson law, accepted.

8. There is one further application which might be mentioned. We might use the index of dispersion to test the null hypothesis that samples have come from a Neyman's contagious distribution with specified m_2. For example, suppose we consider some botanical data provided by Archibald (1948). A plant community was sampled with a quadrat, the number of plants of *Plantago* occurring in each quadrat being noted. It was found that the distribution of the number of quadrats containing 0, 1, 2, ... plants did not follow Poisson's law, and it was suspected that there was clustering of observations owing to the fact that this species is root propagated. In fact, a large number of observations was collected and Neyman's contagious distribution was found to fit the data very well. If now further samples are taken, either from a different locality or from the same locality in a different year, it is interesting to ask whether the variance/mean ratio has changed. For *Plantago*, twenty samples were taken at a different time of year and we have the following quantities:

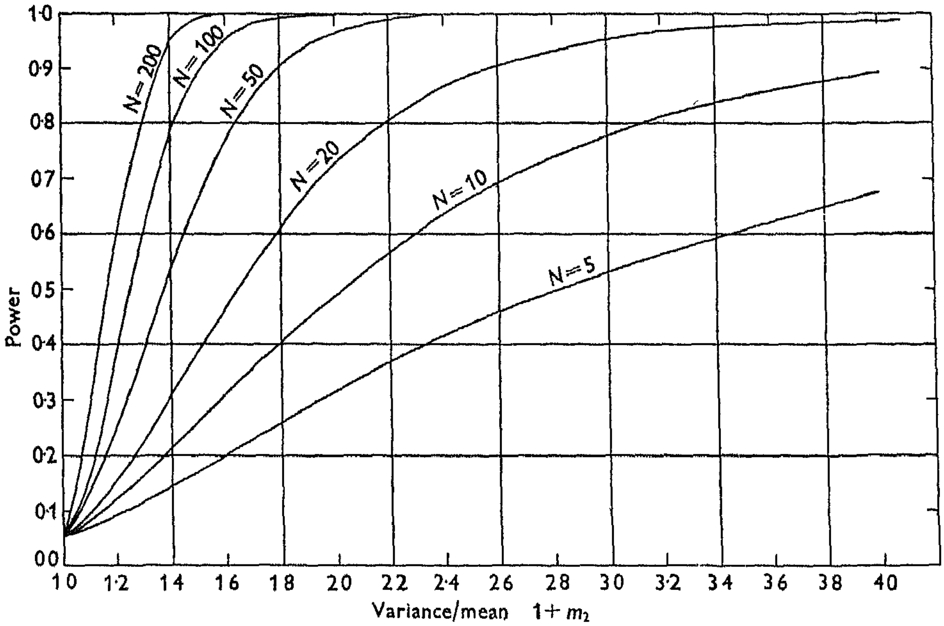

Fig. 1

$$N = 20,\quad \bar{x} = 20{\cdot}5,\quad \sum_{i=1}^{20}(x_i-\bar{x})^2 = 1285{\cdot}0,\quad z_0 = 62{\cdot}68;$$

from previous large-scale work it was found that $m_2 = 2$. Hence for 19 degrees of freedom

$$P\{\chi^2 > z_0/(1+m_2) = 20{\cdot}89\} = 0{\cdot}34.$$

We should conclude from these figures that the variance/mean ratio in the population could have remained unchanged, even though the population mean itself might have altered.

REFERENCES

Archibald, E. E. A. (1948). *Ann. Bot., Lond.*, N.S., **12**, 221.
Hoel, P. G. (1943). *Ann. Math. Statist.* **14**, 155.
Neyman, J. (1939). *Ann. Math. Statist.* **10**, 35.

POWER OF THE MODIFIED t-TEST (u-TEST) BASED ON RANGE

By E. LORD, *Shirley Institute, Didsbury, Manchester*

1. Introduction

In a previous paper (Lord, 1947) we have indicated how estimates of standard deviation based on a total range, or on a mean range, may be used in the so-called u-test when considering the statistical significance of either the deviation of a sample mean from some fixed or hypothetical population value, or the difference between the means of two samples. From tables at several probability levels, rapid estimates of the confidence limits within which the population mean may be assumed to be contained may also be obtained from a measured sample mean and the range or mean range among the individuals. It was pointed out that these applications of the range, in place of root-mean-square estimates of standard deviation in the t-test, necessarily entailed some loss of precision. However, on the basis of some of the results now given in full in the present paper, it was asserted that the reduction in accuracy is negligible for all practical purposes, and is often compensated by the greater ease of computation compared with the usual t-test.

Whilst the original paper was in the press, a note by Daly (1946) was published in which a range test was suggested similar to one of the forms of the u-test. The case when the estimate of standard deviation is made from the total range in a single set of n values was discussed, and approximate values of the ratio (deviation)/(range) on the 10% probability level were given for several low values of n. For two particular values, the power of the range test was found comparable with the power of the conventional t-test. Walsh (1947) also considered forms of a significance test based on a single range ($n \leqslant 10$) and suggested that, for samples of this size, discrimination is nearly as powerful as for the t-test.

In various forms, the more general test based on the distribution of

$$u = \frac{\text{normal deviate}}{\text{independent range or mean-range estimate of standard error}}$$

has been found useful in experimental laboratory work for considering significance of differences between sample means and in estimating the accuracy of measures of average character. In the industrial field, its simple nature permits easy routine application to the detection of changes in average quality of processed batches of material and to determining quickly whether averages of samples, tested with a view to purchase or further use, deviate significantly from specified requirements. The present paper is therefore intended to justify more firmly the value of the u-test. In particular, it shows that, over a wide field of useful application, the power to detect the existence of differences is little less than that of the most powerful t-test.

Use of the quantity 'standardized error' in preference to the power function often facilitates comparison of the two tests. It is shown, for example, that if the t-test detects a given real difference 19 times out of 20 when using the 5% probability level, then with the u-test this difference does not need to be more than 2½% greater for it to be detected with equal frequency. Moreover, in many cases the real difference need only be 1% greater for the u-test than for the t-test. Thus for most practical purposes the difference in discrimination is of negligible importance.

When applying the u-test to medium and large samples, there is often a choice of size of possible subgroups into which the observations may be divided to form mean range estimates of standard deviation. Tables are given to indicate which grouping combination in a fixed total number of observations leads to the greatest power of the u-test. It can be shown that, as would be expected, the grouping combination giving greatest power is the same as that yielding the most efficient estimate of standard deviation. When the number of observations permit, the size of the subgroups should generally be $n = 8$, and in any case between $n = 6$ and $n = 10$. Where the whole sample or samples may be divided in several ways into subgroups of size 6–10, the choice of any one in preference to another does not, however, result in any appreciable loss in accuracy.

2. THE POWER OF THE t-TEST

To obtain results with which to compare the power of the u-test, it was first necessary to establish some new values for the power function of the t-test. If x is a variable normally distributed about a mean of zero, with standard deviation σ, and if s is an independent estimate of σ based on ν degrees of freedom having the distribution

$$p(s) = \frac{\nu^{\frac{1}{2}\nu}}{2^{\frac{1}{2}\nu-1}\,\Gamma(\frac{1}{2}\nu)\,\sigma^{\nu}}\, s^{\nu-1}\, e^{-\frac{1}{2}\nu s^2/\sigma^2}, \tag{1}$$

then the ratio $t = x/s$ follows the well-known 'Student' distribution. The t-test may be used either in the symmetrical or asymmetrical form, according to whether we are seeking to detect changes in population mean in both directions or in one only. In the former case, using the 100α % significance level, we reject the null hypothesis that the population mean $\xi = 0$, if $|t| \geqslant t_{\frac{1}{2}\alpha}$, where

$$\alpha = P\{|t| \geqslant t_{\frac{1}{2}\alpha}\} = 2\int_{t_{\frac{1}{2}\alpha}}^{\infty} \frac{\Gamma[\frac{1}{2}(\nu+1)]}{\Gamma(\frac{1}{2}\nu)\sqrt{(\pi\nu)}}\,(1+t^2/\nu)^{-\frac{1}{2}(\nu+1)}\,dt. \tag{2}$$

The power function of the test gives the chance of establishing significance when $\xi \neq 0$. For a given α and ν, the power is a function of $\rho = \xi/\sigma$ and may be written

$$\text{Power} = \beta(\rho) = \beta'(\rho) + \beta''(\rho), \tag{3}$$

where $\beta'(\rho)$ and $\beta''(\rho)$ are the probabilities that, when $\rho > 0$, $t \geqslant t_{\frac{1}{2}\alpha}$ and $t \leqslant -t_{\frac{1}{2}\alpha}$ respectively.

Using the asymmetrical test, values of $\beta'(\rho)$ were first given by Neyman, Iwaszkiewicz & Kolodziejczyk (1935) for $\alpha = 0{\cdot}05$ and $\alpha = 0{\cdot}01$. In the present paper, using the symmetrical test, we have taken $\frac{1}{2}\alpha = 0{\cdot}025$ and $0{\cdot}005$.*

(a) *Computation of $\beta'(\rho)$ and $\beta''(\rho)$ for $\alpha = 0{\cdot}05$*

Putting $\sigma = 1$ for convenience in calculation, we may write

$$\beta'(\rho) = \int_0^{\infty} \left\{\frac{1}{\sqrt{(2\pi)}}\, e^{-\frac{1}{2}(x-\rho)^2} \int_0^{x/t_{\frac{1}{2}\alpha}} p(s)\,ds\right\} dx, \tag{4}$$

where $p(s)$ is given in (1). The expression for $\beta''(\rho)$ is obtained from (4) by writing $-\rho$ for ρ. Except in the case $\nu = 1, 2$, where the integrals assume simple forms, computation was carried out by expressing (4) in terms of the Incomplete Gamma Function integral, $I(u, p)$. Writing

$$u = \sqrt{(\tfrac{1}{2}\nu)}\, x^2/t^2_{\frac{1}{2}\alpha}, \quad p = \tfrac{1}{2}\nu - 1$$

---

* Some results, presented graphically, for the symmetrical test with $\frac{1}{2}\alpha = 0{\cdot}025$ have been given by Ferris, Grubbs & Weaver (1946).

we have from (4)

$$\beta'(\rho) = \int_0^\infty \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}(x-\rho)^2} I(u, \nu)\, dx. \tag{5}$$

$\beta''(\rho)$ is obtained by writing $(x+\rho)$ for $(x-\rho)$, as before.

Taking $\alpha = 0{\cdot}05$, and using tables of the ordinates of the normal curve and of the incomplete Γ-function in (5), a series of framework values of $\beta'(\rho)$ and $\beta''(\rho)$ were obtained by quadrature methods for $\rho = 1, 2, \ldots, 10$ and $\nu = 1, 2, 3, 4, 5, 9, 14$ and 20. With argument proportional to $1/\nu$ and reciprocals of $\beta'(\rho)$ and $\beta''(\rho)$ as variate, a seven-point Lagrangian formula was applied to points corresponding to $\nu = 3, 4, 5, 9, 14, 20$ and ∞, and interpolation made of the components $\beta'(\rho)$ and $\beta''(\rho)$ for $\nu = 6, 7, 8, 10, 12, 16$ and 18. The interval was then halved using a nine-point Lagrangian formula to obtain the remaining intermediate values. The framework and interpolated values of $\beta'(\rho)$ and $\beta''(\rho)$ are given in Table 1 for integral values of ρ from 1 to 10 and ν from 1 to 20.

The sum $\beta'(\rho)+\beta''(\rho)$ gives the power function of the symmetrical test (using a 5 % significance level) with regard to the alternatives $\rho > 0$ and $\rho < 0$. On the other hand, $\beta'(\rho)$ gives the power function of the asymmetrical test (using a $2\frac{1}{2}$ % significance level) with regard to alternatives $\rho > 0$.

In dealing with the t-test, Neyman & Tokarska (1936) have defined the 'standardized error', ρ_α, as the value of ρ for which

$$1-\beta(\rho) = \alpha.^*$$

Thus, if we use the 5 % significance level, the chance will be 0·95 of establishing significance and 0·05 of failing to do so, if the population mean has shifted from its position under the null hypothesis by an amount $\xi = \rho_{0{\cdot}05}\sigma$. In what follows we shall use $\rho_{0{\cdot}05}$ and $\rho_{0{\cdot}01}$ in order to compare the efficiency of the t- and u-tests.

(b) *Computation of* $\rho_{0{\cdot}05}$ *and* $\rho_{0{\cdot}01}$

For $\alpha = 0{\cdot}05$, approximate values of $\rho_{0{\cdot}05}$ corresponding to $\beta(\rho) = 0{\cdot}95$ were obtained for the framework values of ν by interpolation in Table 1. To the number of figures used in this table it will be noticed that at $\rho = \rho_{0{\cdot}05}$ the value $\beta''(\rho) = 0$, so that $\beta(\rho) = \beta'(\rho)$. Accurate values of $\beta'(\rho)$, and hence of $\beta(\rho)$, were then determined for these framework values of ν by numerical integration of (4) for a range of values of ρ (in steps of 0·005) close to the approximate interpolated values of $\rho_{0{\cdot}05}$. The exact values were determined by subsequent backward interpolation; the value of $\rho_{0{\cdot}05}$ for indefinitely large ν was obtained from integral tables of the normal curve, the limiting form of the t-distribution.

For $\alpha = 0{\cdot}01$, tables of $\beta'(\rho)$ and $\beta''(\rho)$ had not been computed. The values $\rho = \rho_{0{\cdot}01}$ were determined for the framework values of ν in a similar manner, first computing the values of $\beta'(\rho)$ [equal to $\beta(\rho)$ to the 5-figure accuracy used] by quadrature applied to equation (4) and then using backward interpolation. As before, the value of $\rho_{0{\cdot}01}$ for indefinitely large ν was taken from integral tables of the normal curve.

Lagrangian interpolation was then applied to these two sets of framework values to complete the range of values of $\rho_{0{\cdot}05}$ and $\rho_{0{\cdot}01}$ given in Table 2. The values were differenced as a check and are probably correct to one unit in the last decimal place.

---

* In terms of the 'first' and 'second' kinds of errors, this may be written $P_{\text{II}} = P_{\text{I}} = \alpha$.

Table 1. *Power components of the t-test* ($\alpha = 0{\cdot}05$)

	Values of $\beta'(\rho)$										Values of $\beta''(\rho)$		
ν \ ρ	1	2	3	4	5	6	7	8	9	10	1	2	3
1	0·06776	0·12522	0·18611	0·24636	0·30516	0·36218	0·41714	0·46978	0·51989	0·56731	0·00523	0·00053	0·00002
2	0·09132	0·21800	0·38739	0·56457	0·71923	0·83574	0·91284	0·95805	0·98169	0·99275	0·00388	0·00030	0·00001
3	0·10808	0·28854	0·53277	0·75497	0·89860	0·96700	0·99155	0·99830	0·99973	0·99997	0·00319	0·00020	0·00001
4	0·11946	0·33582	0·61927	0·84331	0·95484	0·99096	0·99874	0·99988	0·99999	1·00000	0·00280	0·00015	0·00000
5	0·12742	0·36803	0·67217	0·88726	0·97521	0·99656	0·99970	0·99998	1·00000	—	0·00255	0·00013	—
6	0·13324	0·39093	0·70668	0·91174	0·98413	0·99833	0·99990	0·99999	—	—	0·00238	0·00011	—
7	0·13765	0·40791	0·73060	0·92678	0·98871	0·99903	0·99996	1·00000	—	—	0·00225	0·00009	—
8	0·14110	0·42094	0·74799	0·93673	0·99134	0·99937	0·99998	—	—	—	0·00216	0·00009	—
9	0·14386	0·43123	0·76114	0·94371	0·99300	0·99955	0·99999	—	—	—	0·00209	0·00008	—
10	0·14612	0·43956	0·77139	0·94883	0·99411	0·99966	1·00000	—	—	—	0·00203	0·00007	—
11	0·14800	0·44641	0·77959	0·95273	0·99490	0·99973	—	—	—	—	0·00198	0·00007	—
12	0·14960	0·45214	0·78628	0·95578	0·99548	0·99978	—	—	—	—	0·00195	0·00006	—
13	0·15097	0·45700	0·79184	0·95822	0·99592	0·99981	—	—	—	—	0·00192	0·00006	—
14	0·15215	0·46117	0·79654	0·96022	0·99626	0·99984	—	—	—	—	0·00189	0·00006	—
15	0·15319	0·46479	0·80056	0·96189	0·99653	0·99986	—	—	—	—	0·00186	0·00006	—
16	0·15410	0·46797	0·80402	0·96330	0·99675	0·99987	—	—	—	—	0·00184	0·00006	—
17	0·15491	0·47078	0·80704	0·96450	0·99693	0·99988	—	—	—	—	0·00182	0·00005	—
18	0·15563	0·47328	0·80970	0·96553	0·99709	0·99989	—	—	—	—	0·00181	0·00005	—
19	0·15628	0·47553	0·81206	0·96643	0·99723	0·99990	—	—	—	—	0·00179	0·00005	—
20	0·15687	0·47756	0·81416	0·96722	0·99735	0·99991	—	—	—	—	0·00178	0·00005	—
∞	0·16854	0·51597	0·85084	0·97933	0·99882	0·99997	1·00000	1·00000	1·00000	1·00000	0·00154	0·00004	—

Table 2. *Standardized error of the t-test*

ν	$\rho_{0\cdot05}$	$\rho_{0\cdot01}$	ν	$\rho_{0\cdot05}$	$\rho_{0\cdot01}$
1	24·975+	163·99	13	3·907	5·67
2	7·772	21·49	14	3·883	5·61
3	5·654	11·76	15	3·862	5·55+
4	4·928	8·97			
5	4·574	7·73	16	3·844	5·50
			17	3·829	5·46
6	4·367	7·05−	18	3·816	5·43
7	4·232	6·62	19	3·804	5·40
8	4·137	6·33	20	3·793	5·37
9	4·067	6·13			
10	4·013	5·97	30	3·727	5·20
			60	3·665−	5·04
11	3·970	5·85−	120	3·634	4·97
12	3·936	5·75+	∞	3·605−	4·90

3. THE POWER OF THE u-TEST

In the u-test, range estimates of the population standard deviation are used in place of root-mean-square estimates; thus

$$u = u(m,n) = \frac{x}{\bar{w}(m,n)/d_n} \tag{6}$$

replaces t where, as before, under the hypothesis tested, x is distributed normally about a mean of zero. The quantity $\bar{w} = \bar{w}(m,n)$ is the mean value of m ranges, obtained from m independent samples or subgroups, each containing n observations. The constant d_n* is the expected value of the range in samples of n, selected at random from a normal population of unit standard deviation. The ratio $\bar{w}(m,n)/d_n$ is an estimate of the standard error of x determined from range. For a series of values of m and n, tables of the significance levels of u were given in an earlier paper (Lord, 1947) for a number of values of α.

Although the t-test is the uniformly most powerful unbiased test of Student's hypothesis, (Neyman & Pearson, 1936), for many purposes the u-test has the advantage of simplicity and ease of computation. In section (*a*) below the power of the u-test is investigated for the case where the estimate of the standard error of x is determined from a single range (usually the total range in a sample of n values of which the variate x is the mean). The case of $m \geqslant 2$ is considered later. As in the analysis of the t-test, the investigation is limited to the probability levels $\alpha = 0{\cdot}05$ and $\alpha = 0{\cdot}01$ in the symmetrical case; from this, generalizations may be drawn regarding the $\alpha = 0{\cdot}025$ and $\alpha = 0{\cdot}005$ probability levels of the asymmetrical case.

(*a*) *Computation of power and standardized error for the case* $m = 1$

The analysis follows lines similar to those used in the investigation of the t-test. If $p(w)$ is the probability distribution of the range w in a sample of n observations, then corresponding to equation (4) we have

$$\beta'(\rho) = \int_0^\infty \left\{\frac{1}{\sqrt{(2\pi)}}\, e^{-\frac{1}{2}(x-\rho)^2} \int_0^{x d_n/u_{\frac{1}{2}\alpha}} p(w)\, dw\right\} dx. \tag{7}$$

---

* Values of d_n, computed by L. H. C. Tippett (1925), have been published in many places; see, for example, Lord (1947, p. 67).

The percentage points, $u_{\frac{1}{2}\alpha}$, of the distribution of u are analogous to those, $t_{\frac{1}{2}\alpha}$, of the distribution of t. The expression for $\beta''(\rho)$ is similar to that of (7), obtained by writing $(x+\rho)$ for $(x-\rho)$.

For $n = 2$, the values of $\beta'(\rho)$ and $\beta''(\rho)$ for u are identical with those for t with $\nu = 1$, since there is one-to-one correspondence between the range and root-mean-square estimates of σ. Again for indefinitely large n, the values of the power are the same for t and u. For intermediate values of n, fresh computation is necessary.

For $n \geqslant 3$, the integral $\int_0^{xd_n/u_{\frac{1}{2}\alpha}} p(w)\,dw$ was determined at intervals of 0·25 for x, using manuscript tables of the integral $\int_0^{w} p(w)\,dw$ tabulated to 5 figures and loaned by H. O. Hartley. Using equation (7) with a value of $\alpha = 0{\cdot}05$, $\beta'(\rho)$ and $\beta''(\rho)$ were computed by quadrature methods for values of $\rho = 1, 2, \ldots, 10$ and framework values of $n = 3, 4, 5, 6, 10, 14$ and 20. A seven-point general Lagrangian formula, with argument proportional to $1/n$ and reciprocals of the variate, was applied to the points corresponding to $n = 4, 5, 6, 10, 14, 20$ and ∞ to determine values of $\beta'(\rho)$ and $\beta''(\rho)$ for $n = 7, 8, 12, 16$ and 18. The interval was then halved, using a nine-point Lagrangian, through even values of n from 4 to 20. The framework and interpolated values of the two components of $\beta(\rho)$ are given in Table 3.

Approximate values of the standardized error were determined for framework values of n by interpolation in Table 3. As before, in the region of $\rho = \rho_{0{\cdot}05}$, we have $\beta'(\rho) = \beta(\rho)$ to the limit of accuracy of the tabulated figures. A range of values of $\beta(\rho)$ was obtained by applying quadrature to equation (7) for several values of ρ close to the interpolated approximate values of $\rho_{0{\cdot}05}$, ρ increasing in steps of 0·005. Exact values of $\rho_{0{\cdot}05}$ corresponding to $\beta(\rho) = 0{\cdot}95$ were determined using backward interpolation. Lagrangian interpolation was then applied to these exact framework values of $\rho_{0{\cdot}05}$ to determine the remaining corresponding values in the range of n from 2 to 20.

For $\alpha = 0{\cdot}01$, tables of $\beta'(\rho)$ and $\beta''(\rho)$ were not determined over the field of ρ under consideration and consequently approximations to $\rho_{0{\cdot}01}$ could not be obtained by immediate interpolation as in the case of $\rho_{0{\cdot}05}$ above. Nevertheless, computation proceeded on similar lines: quadrature was applied to equation (7) for the framework values of n to obtain in each case a series of values of $\beta'(\rho) = \beta(\rho)$ corresponding to closely and equally spaced values of ρ. By backward interpolation exact framework values of $\rho_{0{\cdot}01}$ were found, and then Lagrangian interpolation used to complete the remaining values. The two sets of values of the standardized error, $\rho_{0{\cdot}05}$ and $\rho_{0{\cdot}01}$, are given in Table 4 and should not be in error by more than 1 unit in the last figure.

(b) *Computation of standardized error for the case* $m \geqslant 2$

In the generalized form of the u-test the estimate of the population standard error of x is made not from a single range, w, but from the mean of m ranges, $\bar{w}(m, n)$ obtained from m independent samples or subgroups, each containing n observations. To simplify the algebraic expressions in what follows, this mean range $\bar{w}(m, n)$ is written as $\bar{w}$ where no confusion is involved.

For a given probability level α and a mean value of x at a distance $\rho > 0$ from the origin (the standard error of x being taken as the unit), we may write the component $\beta'(\rho)$ of the power of the u-test as

$$\beta'(\rho) = \iint \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}(x-\rho)^2} p(\bar{w})\,dx\,d\bar{w}, \tag{8}$$

Table 3. *Power components of the u-test* ($m = 1, \alpha = 0{\cdot}05$)

	Values of $\beta'(\rho)$										Values of $\beta''(\rho)$		
n \ ρ	1	2	3	4	5	6	7	8	9	10	1	2	3
2	0·06776	0·12512	0·18611	0·24636	0·30516	0·36218	0·41714	0·46978	0·51989	0·56731	0·00523	0·00053	0·00002
3	0·09129	0·21784	0·38689	0·56354	0·71775	0·83409	0·91127	0·95679	0·98082	0·99223	0·00388	0·00030	0·00001
4	0·10784	0·28725	0·52957	0·75032	0·89424	0·96419	0·99026	0·99786	0·99961	0·99994	0·00320	0·00020	0·00001
5	0·11882	0·33269	0·61263	0·83588	0·95003	0·98887	0·99825	0·99980	0·99998	1·00000	0·00281	0·00016	0·00000
6	0·12638	0·36309	0·66283	0·87865	0·97097	0·99537	0·99950	0·99996	1·00000	—	0·00257	0·00013	—
7	0·13176	0·38425	0·69525	0·90270	0·98053	0·99758	0·99979	0·99998	—	—	0·00241	0·00011	—
8	0·13579	0·39985	0·71774	0·91768	0·98560	0·99850	0·99989	0·99999	—	—	0·00229	0·00009	—
9	0·13892	0·41176	0·73411	0·92768	0·98855	0·99896	0·99993	1·00000	—	—	0·00221	0·00009	—
10	0·14141	0·42113	0·74656	0·93488	0·99057	0·99923	0·99995	—	—	—	0·00214	0·00008	—
11	0·14343	0·42866	0·75628	0·94020	0·99196	0·99940	0·99997	—	—	—	0·00208	0·00008	—
12	0·14510	0·43484	0·76405	0·94427	0·99289	0·99951	0·99998	—	—	—	0·00204	0·00007	—
13	0·14652	0·44001	0·77042	0·94750	0·99361	0·99959	0·99998	—	—	—	0·00201	0·00007	—
14	0·14774	0·44441	0·77577	0·95013	0·99419	0·99965	0·99999	—	—	—	0·00198	0·00007	—
15	0·14880	0·44821	0·78031	0·95230	0·99465	0·99970	0·99999	—	—	—	0·00195	0·00007	—
16	0·14973	0·45154	0·78422	0·95413	0·99501	0·99973	0·99999	—	—	—	0·00193	0·00007	—
17	0·15056	0·45449	0·78763	0·95569	0·99531	0·99976	1·00000	—	—	—	0·00191	0·00006	—
18	0·15130	0·45712	0·79066	0·95705	0·99557	0·99978	—	—	—	—	0·00190	0·00006	—
19	0·15197	0·45949	0·79336	0·95824	0·99579	0·99980	—	—	—	—	0·00189	0·00006	—
20	0·15259	0·46164	0·79577	0·95928	0·99598	0·99981	—	—	—	—	0·00187	0·00006	—
∞	0·16854	0·51597	0·85084	0·97933	0·99882	0·99997	1·00000	1·00000	1·00000	1·00000	0·00154	0·00004	—

Table 4. *Standardized error of the u-test* ($m = 1$)

n	$\rho_{0\cdot05}$	$\rho_{0\cdot01}$	n	$\rho_{0\cdot05}$	$\rho_{0\cdot01}$
2	24·975+	163·99	12	4·062	6·11
3	7·807	21·66	13	4·028	6·01
4	5·713	11·98	14	3·999	5·93
5	5·000	9·21	15	3·975−	5·86
6	4·652	7·98	16	3·954	5·80
7	4·451	7·30	17	3·936	5·75−
8	4·320	6·88	18	3·920	5·70
9	4·227	6·59	19	3·906	5·66
10	4·158	6·39	20	3·893	5·62
11	4·105−	6·23	∞	3·605−	4·90

where the limits of integration satisfy

$$0 \leqslant x < \infty,$$

$$0 \leqslant \overline{w} < \infty,$$

$$\frac{x}{\overline{w}/d_n} \geqslant u_{\frac{1}{2}\alpha},$$

and where $y = p(\overline{w})$ denotes the probability distribution of $\overline{w} = \overline{w}(m, n)$. Unlike equation (7), we shall integrate (8) first with respect to x and obtain

$$\beta'(\rho) = \int_0^\infty p(\overline{w})\, d\overline{w} \int_{\overline{w}u_{\frac{1}{2}\alpha}/d_n - \rho}^{\infty} \frac{1}{\sqrt{(2\pi)}}\, e^{-\frac{1}{2}x^2}\, dx.$$

Similarly, it is easy to show that the second component of the power, $\beta''(\rho)$, is a similar expression with $+\rho$ written for $-\rho$ in the limit of integration.

In this section, values of the standardized error $\rho_{0\cdot05}$ are determined over a range of values of m and n. As before, the second component, $\beta''(\rho)$, of the power function is zero in the region of $\rho = \rho_{0\cdot05}$ to the limits of accuracy used (five decimal places or more), and we therefore have $\beta(\rho) = \beta'(\rho)$.

For a series of framework values of m and n, and using numerical values ($u_{\frac{1}{2}\alpha}$) of the 5 % points of the distribution of $u = u(m, n)$ tabulated previously (Lord, 1947), quadrature was applied in each case to equation (8) to obtain a series of values $\beta(\rho)$ corresponding to a range of narrowly spaced values of ρ. By backward interpolation $\rho_{0\cdot05}$ was then determined as the value of ρ corresponding to $\beta(\rho) = 0\cdot95$. This calculation was carried out to obtain framework values corresponding to $m = 2$, $n = 2, 3, 4, 5, 6, 10, 16$ and 20; $m = 4$, $n = 2, 3$ and 4; and $m = 8$ and $n = 2$, using values of the ordinate $p(\overline{w})$ of the distribution of the mean range previously evaluated in the investigation of the u-test cited above. The limiting value for $\rho_{0\cdot05}$ for indefinitely large m or indefinitely large n is, as in the preceding section, the limiting value of $\rho_{0\cdot05}$ for indefinitely large ν in the t-test.

For $m = 2$, Lagrangian formulae were used to interpolate the intermediate values of the standardized error, taking arguments proportional to $1/n$ and reciprocals of the corresponding standardized error for the variate. As a check, the values were inspected by determining the differences up to the second order, and then reduced by one place of decimals. These reduced values are given in Table 5, together for convenience with reduced values for $m = 1$ from

Table 5. *Standardized error of the u-test* ($\alpha = 0{\cdot}05$)

n \ *m*	1	2	3	4	5	6	7	8	9	10	15	20	30	60	120
2	24·98	7·97	5·82	5·06	4·68	4·45+	4·31	4·20	4·12	4·06	3·90	3·82	3·74	3·67	3·64
3	7·81	5·00	4·42	4·17	4·04	3·96	3·90	3·86	3·83	3·81	3·74	3·70	3·67	3·64	3·62
4	5·71	4·41	4·10	3·96	3·88	3·83	3·80	3·77	3·75+	3·74	3·69	3·67	3·65−	3·63	3·62
5	5·00	4·18	3·97	3·87	3·81	3·78	3·75+	3·73	3·72	3·71	3·67	3·65+	3·64	3·62	3·61
6	4·65+	4·05+	3·89	3·81	3·77	3·74	3·72	3·71	3·69	3·69	3·66	3·65−	3·63	3·62	3·61
7	4·45+	3·98	3·84	3·78	3·74	3·72	3·70	3·69	3·68	3·67	3·65−	3·64	3·63	3·62	3·61
8	4·32	3·92	3·81	3·76	3·73	3·70	3·69	3·68	3·67	3·66	3·64	3·63	3·62	3·61	3·61
9	4·23	3·89	3·79	3·74	3·71	3·69	3·68	3·67	3·66	3·66	3·64	3·63	3·62	3·61	3·61
10	4·16	3·86	3·77	3·73	3·70	3·68	3·67	3·66	3·66	3·65+	3·64	3·63	3·62	3·61	3·61
11	4·10	3·84	3·76	3·72	3·69	3·68	3·67	3·66	3·65+	3·65−	3·63	3·63	3·62	3·61	3·61
12	4·06	3·82	3·74	3·71	3·69	3·67	3·66	3·66	3·65−	3·66−	3·63	3·62	3·62	3·61	3·61
13	4·03	3·80	3·73	3·70	3·68	3·67	3·66	3·65+	3·65−	3·64	3·63	3·62	3·62	3·61	3·61
14	4·00	3·79	3·73	3·69	3·68	3·66	3·66	3·65−	3·64	3·64	3·63	3·62	3·62	3·61	3·61
15	3·97	3·78	3·72	3·69	3·67	3·66	3·65+	3·65−	3·64	3·64	3·63	3·62	3·62	3·61	3·61
16	3·95+	3·77	3·71	3·69	3·67	3·66	3·65+	3·64	3·64	3·64	3·63	3·62	3·62	3·61	3·61
17	3·94	3·76	3·71	3·68	3·67	3·66	3·65−	3·64	3·64	3·64	3·62	3·62	3·61	3·61	3·61
18	3·92	3·75+	3·70	3·68	3·66	3·65+	3·65−	3·64	3·64	3·63	3·62	3·62	3·61	3·61	3·61
19	3·91	3·75−	3·70	3·67	3·66	3·65+	3·64	3·64	3·64	3·63	3·62	3·62	3·61	3·61	3·61
20	3·89	3·74	3·70	3·67	3·66	3·65−	3·64	3·64	3·63	3·63	3·62	3·62	3·61	3·61	3·61

Table 4. In the case of $n = 2$, framework values of $\rho_{0\cdot05}$ had been calculated for $m = 4$ and $m = 8$. Values of $\rho_{0\cdot05}$ were determined for each chosen value of m by Lagrangian interpolation, using reciprocals of m for argument and reciprocals of the corresponding values of $\rho_{0\cdot05}$ as the variate, the curve passing through the points corresponding to $n = 1, 2, 4, 8$ and ∞. The interpolated values of $\rho_{0\cdot05}$ obtained by this method, and the directly computed values for $m = 4$ and $m = 8$ are given in Table 5 for a series of values of m suitable for practical use. As a check upon the accuracy of this method of interpolation the computation was repeated, this time using a four-point Lagrangian passing through points corresponding to $m = 1, 2, 4$ and ∞. Most of the interpolated four-point values of $\rho_{0\cdot05}$ agree exactly with those obtained by five-point interpolation. Where differences occur, none amount to more than either 1 or rarely 2 units in the last tabulated figure. The five-point Lagrangian method of interpolation therefore appears adequate for furnishing the required degree of accuracy, an error not exceeding 1 unit in the last figure.

For $n = 3$ and $n = 4$ the values of the standardized error for different values of m are appreciably smaller and are nearer to their limiting values than in the case of $n = 2$. In view of the agreement between the four-point and five-point Lagrangian interpolated values of $\rho_{0\cdot05}$ for $n = 2$, a four-point Lagrangian through points corresponding to $m = 1, 2, 4$ and ∞ is adequate for interpolation of $\rho_{0\cdot05}$ for these further two values of n.

As a further check upon the validity of the above method, the values of $\rho_{0\cdot05}$ for $n = 4$ were also interpolated for the chosen range of values of m, using a three-point Lagrangian curve through points corresponding to $m = 1, 2$ and ∞. This yielded interpolated values in exact agreement (to the number of figures given) with the four-point values in Table 5. The accuracy of the three-point Lagrangian will certainly be no less for interpolating within the field of the less rapidly changing values of $\rho_{0\cdot05}$. It was applied in the case of values of n from 5 to 20 to obtain interpolated values of the standardized error corresponding to the chosen range of values of m, and the results, after reduction to two decimal places, are given in Table 5.

4. EFFECT OF GROUPING IN THE u-TEST

The case when the estimate of standard error of x is obtained from the mean of m ranges of n random values, instead of a single range, has been investigated for the $\alpha = 0\cdot05$ probability level by the determination of the standardized error $\rho_{0\cdot05}$. Reference to Table 5 shows that for a sample containing, say, $N = mn$ observations the standardized error is not constant but varies slightly for the different possible combinations of m and n. For example, taking $N = 16$ we have the following values of $\rho_{0\cdot05}$: 4·20 for $m = 8$ and $n = 2$; 3·96 for $m = n = 4$; 3·92 for $m = 2$ and $n = 8$; and 3·95 for $m = 1$ and $n = 16$.

The variation in the values of $\rho_{0\cdot05}$ is more pronounced for small samples than for large samples. For $N = 10$, $\rho_{0\cdot05}$ falls from the value 4·68, corresponding to $m = 5$ and $n = 2$, to 4·16 when the estimate of standard deviation is determined from the total range of values, i.e. when $m = 1$ and $n = N = 10$. For larger samples $\rho_{0\cdot05}$ falls from a maximum at $n = 2$ to a minimum for a particular combination of m and n, and then increases again as n approaches the sample size N. When the size of the sample is such that it may be divided in several ways into random subgroups of different size n, the greatest power of discrimination corresponding to the minimum standardized error occurs when $n = 8$. As would be expected, it can easily be shown that this size of subgroup also gives the most efficient mean-range estimate of standard deviation of indefinitely large samples. More particularly, comparison of the values of $\rho_{0\cdot05}$ in Table 5 with the corresponding figures for the standard error of mean-range estimates of standard deviation of samples of finite size $N = mn$, indicates that the grouping combination giving the greatest power of discrimination in the u-test is the same as that yielding the most accurate estimates of standard deviation.

The loss of power of the u-test, attendant upon the use of a number of subgroups of very small size, is caused primarily because we ignore the variation between the subgroups in forming an estimate of the variate standard error. A similar loss of power also arises in the application of the usual t-test, if the estimate of standard deviation is determined from the variation in subgroups of a sample of total size N. Here the loss is represented by a reduction in the degrees of freedom of the estimate of variance, from $N-1$ to $N-m$.

If N be a prime number, it is still preferable to divide the sample into subgroups for estimating the standard deviation by mean range. The theory indicates that however large N, the u-test is nearly as powerful as the t-test. Nevertheless, in practice it is advisable not to overlook the possibility that the variate does not exactly follow the normal law, particularly at the tails of the distribution where the extreme individuals from large samples will fall. Again, N may not be prime but contain only one low factor or only large factors. If all the observations were used, the estimate of standard deviation would be determined from the mean range in either a large number of very small subgroups (of size, say, $n = 2$ or 3) or a small number of very large subgroups. In both instances, it is generally advisable to ignore a few observations and divide the major portion of the sample into random subgroups of equal size. Two illustrative examples bearing on these points are given below.

Suppose we wish to determine whether the mean ($\bar{x}$) of a sample of 19 values is consistent with the parent population having some hypothetical mean value ξ. To test this null hypothesis we could determine the ratio

$$u = \frac{|\bar{x}-\xi|\, d_{19}\sqrt{(19)}}{w_{19}}.$$

With u based on the total range, $m = 1$ and $n = 19$, Table 5 gives $\rho_{0\cdot05} = 3\cdot91$. A slightly greater degree of accuracy would result from the use of an estimate of standard deviation based on $m = 2$ ranges of $n = 9$ random values, for which $\rho_{0\cdot05} = 3\cdot89$. It is of course essential that the choice of the groupings of 9 should be random. If the order of occurrence of the observations is known to be independent of their magnitude, as frequently happens in practice, we would probably accept the first nine numbers to form one group, the 10th to the 18th as the second group, and ignore the 19th in determining the range estimate of standard deviation. The ignored value is, of course, included with the others in computing the sample mean $\bar{x}$, and the corresponding u would be $|\bar{x}-\xi|\, d_9\sqrt{(19)}/\bar{w}(2,9)$.

As a further illustration of the effect of grouping, we shall consider the case when the mean ($\bar{x}_1$) of a sample of say 34 values is to be compared with the mean ($\bar{x}_2$) of a second sample containing 51 values. We could estimate the standard error of the difference between $\bar{x}_1$ and $\bar{x}_2$ from the mean $\bar{w}(5,17)$ of 5 ranges, each of 17 values, i.e. using 2 random subgroups of 17 values from the first sample and 3 random subgroups of 17 from the second. The value of u would be

$$u = \frac{|\bar{x}_1-\bar{x}_2|\, d_{17}}{\bar{w}(5,17)\sqrt{(\frac{1}{34}+\frac{1}{51})}},$$

and its statistical significance could be gauged by comparing its value with tabulated percentage points of $u(5,17)$. From Table 5, the corresponding standardized error of the u-test for $m = 5$ and $n = 17$ is $3\cdot67$. Alternatively, we might estimate the standard error of the difference from the mean of $m = 8$ ranges of $n = 10$ values, 3 ranges from the first sample (with 4 values discarded at random) and 5 ranges from the second (1 value discarded at random). The appropriate equation for u would then be

$$u = \frac{|\bar{x}_1 - \bar{x}_2|\, d_{10}}{\bar{w}(8, 10)\sqrt{(\frac{1}{34} + \frac{1}{51})}}.$$

By varying the size of subgroups and discarding a few odd values, several other combinations are feasible. The table below contains some combinations which might possibly be chosen in practice with these data, and gives the standardized error $\rho_{0\cdot 05}$ when the range estimate of standard deviation is based on m_1 ranges of n values from the first sample and m_2 from the second. (Values of the standardized error are given to 3 decimal places, calculated from the original values of Table 5 before reduction to 2 decimals. These interpolated values may be in error by 1, less likely by 2, units in the last place.) It will be seen that the efficiency of the test, as measured by $\rho_{0\cdot 05}$, scarcely differs throughout the range of n considered.

n	m_1	m_2	$\rho_{0\cdot 05}$	Observations discarded
17	2	3	3·666	0
10	3	5	3·664	5
9	3	5	3·670	13
8	4	6	3·664	5
7	4	7	3·666	8
6	5	8	3·667	7
5	6	10	3·668	5

Table 5 may be used in cases such as those above, to determine quickly the grouping combination which will give the greatest accuracy in application of the u-test. Generally, it would appear from Table 5 that samples of size $N \geqslant 13$ should be divided into subgroups of equal size and the estimate of standard error made from the mean range within the subgroups. Where possible, the size of the subgroups should be $n = 8$, in any case not greater than $n = 10$ nor less than $n = 6$. The division of a whole sample in this way may entail discarding a few 'surplus' observations in determining a mean range, but these are not excluded from the sample mean. When, as above, the sample may be divided in several ways in subgroups of size from 6 to 10, the choice of any one in preference to another will certainly not lead to any appreciable loss of accuracy.

5. POWER COMPARISON OF t- AND u-TESTS

A comparison of Tables 1 and 3 shows the relative efficiency of the symmetrical t- and u-tests in establishing significance at the 5 % probability level, in the case where the range estimate of σ is derived from a single group of n observations. It is clear that while for a given value of $\rho = \xi/\sigma$ the t-test is always the more powerful, the difference is barely appreciable for many practical purposes.

The power of discrimination of the two tests may also be compared in terms of the standardized errors $\rho_{0\cdot 05}$ and $\rho_{0\cdot 01}$ given in Tables 3 and 4. Table 6 shows the ratio of the standardized error of the u-test to that of the t-test. The maximum value of this ratio for $\rho_{0\cdot 05}$ is seen to be 1·024. This, for example, means that with a sample of $n = 16$, whereas the odds are 19 to 1 on establishing significance at the 5 % level with the t-test if there is a shift in population mean of $\pm 3{\cdot}862\sigma/\sqrt{(16)} = \pm 0{\cdot}9655\sigma$ from the value specified by the null hypothesis, the shift must be nearly $2\frac{1}{2}$ % greater for it to be detected with the same odds if σ is estimated from the range. If the 1 % level is used, and odds of 99 to 1 on establishing significance are taken, a $4\frac{1}{2}$ % greater shift in mean is needed with the u-test than with the t-test.

Standardized errors are particularly applicable to study of the u-test when the estimate of standard error of the variate is based upon the mean of ranges within random subgroups of equal size. For example, suppose we have a single sample of $n = 20$ and wish to establish whether the population mean has a specified value. With the t-test for $\nu = 19$ degrees of freedom, Table 2 gives $\rho_{0\cdot05} = 3{\cdot}804$. With the u-test and using the total range in the 20 values, Table 4 gives $\rho_{0\cdot05} = 3{\cdot}893$. Correspondingly, the ratio $3{\cdot}893/3{\cdot}804 = 1{\cdot}023$ as noted in Table 6. If in place of the total range the mean of two ranges of 10 values be used in the u-test, Table 5 gives 3·86 for $\rho_{0\cdot05}$. This latter value is not quite 2 % greater than $\rho_{0\cdot05}$ for the t-test and is slightly less than the value arising from the total range.

Table 6. *Ratios of standardized errors of u- and t-tests*

n	Ratio of standardized errors		n	Ratio of standardized errors	
	0·05	0·01		0·05	0·01
2	1·000	1·000	12	1·023	1·045 −
3	1·005 −	1·008	13	1·023	1·045 +
4	1·010	1·018	14	1·024	1·045 +
5	1·015 −	1·028	15	1·024	1·045 −
6	1·017	1·033	16	1·024	1·045 −
7	1·019	1·036	17	1·024	1·044
8	1·021	1·039	18	1·024	1·043
9	1·022	1·041	19	1·024	1·043
10	1·022	1·043	20	1·023	1·042
11	1·023	1·044	∞	1·000	1·000

As a further illustration of this method of comparing the tests, we may consider the testing of the statistical significance of the difference between the observed mean values of two samples, each of size, say, $n = 30$. From Table 2, use of the t-test on the 5 % probability level is seen to give a standardized error of about 3·66 ($\nu = 58$ degrees of freedom). With the u-test, if each sample be divided into, say, 3 subgroups of 10 values and the mean of the 6 ranges be used to estimate the standard error of the observed difference, reference to Table 5 gives $\rho_{0\cdot05} = 3{\cdot}68$. This value differs from the t-test value of 3·66 by less than 1 %.

The foregoing investigations appear to justify previous claims made for the use of range in place of root-mean-square estimates of standard deviation in an analogue of the usual t-test. For most practical purposes the loss in power accompanying the use of the range is negligible and, indeed, is generally compensated by the appreciably simpler and quicker method of its application in a wide range of problems frequently arising in both routine testing and investigational work.

REFERENCES

DALY, J. F. (1946). *Ann. Math. Statist.* **17**, 71.
FERRIS, C. D., GRUBBS, F. E. & WEAVER, C. L. (1946). *Ann. Math. Statist.* **17**, 178.
LORD, E. (1947). *Biometrika*, **34**, 41.
NEYMAN, J., IWASZKIEWICZ, K. & KOLODZIEJCZYK, ST. (1935). *J. R. Statist. Soc. Suppl.* **2**, 107.
NEYMAN, J. & PEARSON, E. S. (1936). *Statist. Res. Mem.* **1**, 1.
NEYMAN, J. & TOKARSKA, B. (1936). *J. Amer. Statist. Ass.* **31**, 318.
TIPPETT, L. H. C. (1925). *Biometrika*, **17**, 364.
WALSH, J. E. (1947). *Ann. Math. Statist.* **18**, 280.

THE USE OF MEAN RANGE AS AN ESTIMATOR OF VARIANCE IN STATISTICAL TESTS

By P. B. PATNAIK

1. Introduction

The distribution of the range in random samples from a normal population has been studied numerically by Tippett (1925) and Pearson (1926, 1932). McKay & Pearson (1933) have obtained the general expression for the distribution. Hartley (1942) determined a simpler expression for the probability integral and with Pearson (1942) tabulated this for samples of size $n = 2$ to 20.

Only the first two moments of the range* have so far been fully calculated and are available numerically (Table A, Pearson, 1932), although values of β_1, β_2 are known exactly for n up to 6 and, approximately, after 6 (Pearson, 1926, p. 192). It is possible, however, to compute tables of the exact moments from those of the probability integral by quadrature.†

The use of the range in place of the root-mean-square estimate of the population standard deviation has become important in several fields, particularly in quality control. Although the range furnishes a less efficient estimate, its simplicity and ease of application in routine work, such as production-line inspection, makes it an important statistic in practice.

It is known that the distribution of the range in normal samples depends on the sample size n, is independent of the population mean, but dependent on its standard deviation, σ. Moreover, for large n (say $n > 20$), even small departures from normality in the tails of the parental distribution have a considerable effect on the distribution of range. Furthermore, the relative efficiency of the range as an estimator of σ decreases as n increases. Hence, in practice, when n exceeds say 12, it is preferable to divide the sample into a number of groups and take a weighted mean of the several group ranges. The question of the subdivision which will give the smallest variance of the estimate has been considered by Pearson (1932) and by Grubbs & Weaver (1947). It is found that groups of seven or eight are the best. It will generally be advantageous to have equal groups of about this size and use the average of the group ranges, viz. the mean range.

The exact distribution of mean range cannot be obtained in a simple form, except in the trivial case of two samples of two observations in each. We shall, however, obtain an approximation to the distribution and, from this, derive the distribution of the ratio of a normal variate to mean range and compare it with the results of Lord (1947), who has studied the position numerically. We will also find an approximate distribution of the ratio of an individual range to mean range. These distributions will be used to obtain certain statistical tests.

2. Approximation to the distribution of mean range

Denote by $x_1, x_2, \ldots, x_n$ a random sample of n observations, arranged in ascending order of magnitude, drawn from a normal population having mean μ and standard deviation σ. Then the range in this sample is denoted by

$$w_n = x_n - x_1.$$

---

[*] 'Range' will denote, throughout, the distance between the highest and the lowest observation in a sample randomly selected from a normal population.

[†] [Calculations, for sample sizes up to 20, have now been completed. Ed.]

If there are m such independent samples, with n observations in each, the mean of the m ranges will then be denoted by $\overline{w}_{m,n}$.

Suppose the mean and variance of w_n are written as σd_n and var (w_n). Then the mean and variance of the distribution of $\overline{w}_{m,n}/\sigma$ are given by

$$\left.\begin{aligned} \mathscr{E}(\overline{w}_{m,n}/\sigma) &= d_n = M, \quad \text{say}, \\ \operatorname{var}(\overline{w}_{m,n}/\sigma) &= \frac{1}{m\sigma^2}\operatorname{var}(w_n) = V, \quad \text{say}. \end{aligned}\right\} \tag{1}$$

Hence we see that $\overline{w}_{m,n}/d_n$ gives an unbiased estimate of σ.

Pearson (1926, p. 191) has given a diagram showing the β_1, β_2 points of the distribution of w_n. If the sample contains less than about twelve observations, these points lie in the Type I area and not far from the Type III line; the β_1, β_2 points for the mean range, $\overline{w}_{m,n}$, will lie in the same area, but closer to the normal point, $\beta_1 = 0$, $\beta_2 = 3$. Since the β_1, β_2 points for the distribution of χ fall in the same region and converge, as the degrees of freedom increase, on the normal point, it seems likely that a χ-distribution will give a reasonably accurate representation of the distribution of $\overline{w}_{m,n}$.

Identifying the first two moments, M and V, of $\overline{w}_{m,n}/\sigma$ with those of $c\chi/\sqrt{\nu}$, where χ has ν degrees of freedom, we obtain the following relations:

$$M = \frac{c}{\sqrt{\nu}}\sqrt{2}\,\Gamma\left(\frac{\nu+1}{2}\right)\Big/\Gamma\left(\frac{\nu}{2}\right),$$

$$V = \frac{c^2}{\nu}\left[\nu - 2\left\{\Gamma\left(\frac{\nu+1}{2}\right)\Big/\Gamma\left(\frac{\nu}{2}\right)\right\}^2\right].$$

Expanding the Γ-functions by Stirling's formula, we find

$$M = c\left(1 - \frac{1}{4\nu} + \frac{1}{32\nu^2} + \frac{5}{128\nu^3} + \cdot\right), \tag{2}$$

$$V = c^2\left\{1 - \left(1 - \frac{1}{4\nu} + \frac{1}{32\nu^2} + \frac{5}{128\nu^3} + \ldots\right)^2\right\}, \tag{3}$$

from which follows the equation

$$\frac{V}{M^2} = 1\Big/\left(1 - \frac{1}{2\nu} + \frac{1}{8\nu^2} + \frac{1}{16\nu^3} + \ldots\right) - 1,$$

i.e.
$$k = \frac{V}{M^2} = \frac{1}{2\nu} + \frac{1}{8\nu^2} - \frac{1}{16\nu^3} + \ldots. \tag{4}$$

Retaining the first two terms on the right-hand side of (4) and solving the resulting quadratic, we have

$$1/\nu = -2 + 2\sqrt{(1+2k)}.$$

A further approximation, using the term in $1/(\nu^3)$ gives

$$1/\nu = -2 + 2\sqrt{(1+2j)}, \tag{5}$$

where $j = k + 1/(16\nu^3) = V/M^2 + 1/(16\nu^3)$, the first approximation to ν being used in the second term. The corresponding expression for c, to the same order of approximation as ν in (5), is

$$c = M\left\{1 + \frac{1}{4\nu} + \frac{1}{32\nu^2} - \frac{5}{128\nu^3}\right\}, \tag{6}$$

in which M can be replaced by d_n.

Thus we can represent the distribution of $\bar{w}_{m,n}/(c\sigma)$ by that of $\chi/\sqrt{\nu}$. Since s/σ follows the distribution of $\chi/\sqrt{\nu}$, s being the sample root-mean-square estimate of σ based on ν degrees of freedom, it follows that $\bar{w}_{m,n}$ is distributed approximately as cs.

From the mean and standard deviation of range given in Pearson's Table A (1932), the values of M and V in (1) are known, and using these in (5), ν is obtained and then c from (6). We give here (Table 1) a short table of d_n, ν and c for $n = 3$ to 10, $m = 1$ to 5.

Table 1. *Values of d_n, ν and c for the χ approximation to $w_{m,n}$*

Size of sample n	d_n	No. of samples									
		$m = 1$		$m = 2$		$m = 3$		$m = 4$		$m = 5$	
		ν	c	ν	c	ν	c	ν	c	ν	c
3	1·6926	1·934	1·9164	3·850	1·8049	5·674	1·7684	7·499	1·7498	9·312	1·7385
4	2·0588	2·951	2·2374	5·705	2·1505	8·450	2·1204	11·191	2·1052	13·931	2·0960
5	2·3259	3·828	2·4812	7·474	2·4048	11·107	2·3788	14·732	2·3657	18·355	2·3578
6	2·5344	4·692	2·6721	9·169	2·6001	13·641	2·5812	18·109	2·5696	22·576	2·5626
7	2·7044	5·499	2·8295	10·779	2·7677	16·052	2·7468	21·324	2·7362	26·596	2·7299
8	2·8472	6·259	2·9630	12·297	2·9056	18·328	2·8850	24·358	2·8768	30·386	2·8707
9	2·9700	6·989	3·0778	13·753	3·0245	20·511	3·0064	27·270	2·9974	34·025	2·9920
10	3·0775	7·689	3·1789	15·151	3·1287	22·609	3·1117	30·066	3·1032	37·523	3·0981

N.B. $w_{m,n}/\sigma$ is distributed approximately as $c\chi/\sqrt{\nu}$, where c is a scale factor and ν a fractional degree of freedom.

Unfortunately, there is at present no direct means of judging the accuracy of the above approximation. When the exact moments of w_n above the second are computed, it will be possible to derive the higher moments of $\bar{w}_{m,n}$ and so compare them with the moments of our approximate s-distribution. The probability integral and percentage points of the distribution of $\bar{w}_{m,n}$ can be obtained approximately from those of the χ and hence from those of the χ^2-distribution with degrees of freedom given by the ν of equation (5).

Turning to Table 1, we note the way in which c tends to d_n from above as n increases; further, we see that the equivalent, 'fractional' degrees of freedom, ν, of the approximation are always rather less than $m(n-1)$, the figure appropriate were we to estimate σ from the within-group sums of squares.

3. DISTRIBUTION OF $x/\bar{w}$

Let x be a variate distributed normally with mean zero and standard deviation σ, and let s be an independent root-mean-square estimate of σ based on ν degrees of freedom. In § 2 it was seen that the mean range could be used for estimating σ. Therefore in parallel with the ratio t, we may take the ratio

$$u = \frac{x}{\bar{w}/d_n}, \tag{7}$$

where $\bar{w}$ stands for $\bar{w}_{m,n}$[*] and $d_n\sigma$ is the expectation of the range in samples of n.

Lord (1947, 1950) has considered this distribution and developed the 'u-test', giving several of its applications. A wide range of tables of percentage points have been computed by him by methods of quadrature and interpolation. Daly (1946) also deals with this test, but he takes a single range instead of the mean range.

---

[*] In the rest of this section $\bar{w}$ will be used for $\bar{w}_{m,n}$.

We will now employ the approximate distribution of $\bar{w}$ obtained in § 2 to derive an approximation to the distribution of u. Since $\bar{w}$ is represented by cs, we see that the distribution of u can be approximated by that of

$$u = \frac{d_n x}{cs} = \frac{d_n}{c} t, \tag{8}$$

where t is based on ν degrees of freedom, c and ν being given by equations (5) and (6). In other words, $u\{1 + 1/(4\nu) + 1/(32\nu^2) - 5/(128\nu^3)\}$ is distributed approximately as 'Student's' t with degrees of freedom ν given by equation (5). It is therefore possible to find approximate percentage points for u by interpolating for a fractional ν in any table of the percentage points of t. Using this method, we have calculated the percentage points of the distribution for significance levels* $\alpha = 0{\cdot}10$, $0{\cdot}05$, $0{\cdot}02$ and $0{\cdot}01$ for several values of n and m. They are shown in Table 2, and Lord's values are given alongside for comparison. The good agreement, even for low values of α and for $m = 1$, i.e. for a single range, suggests that the distribution of u is closely approximated by that of $d_n t/c$. We may also infer that the χ-approximation to the distribution of $\bar{w}$ is fairly good.

Table 2. *Values of the percentage points of the distribution of* $u = d_n x/\bar{w}_{m,n}$

Size of sample n	No. of samples m	$\alpha = 0{\cdot}10$		$\alpha = 0{\cdot}05$		$\alpha = 0{\cdot}02$		$\alpha = 0{\cdot}01$	
		Approx.*	Lord†	Approx.*	Lord†	Approx.*	Lord†	Approx.*	Lord†
5	1	2·02	2·02	2·64	2·63	3·58	3·56	4·39	4·38
	3	1·75	1·75	2·15	2·15	2·65	2·65	3·03	3·02
	5	1·71	1·71	2·07	2·07	2·51	2·51	2·83	2·83
	10	1·68	1·68	2·01	2·01	2·42	2·42	2·70	2·70
8	1	1·85	1·85	2·33	2·32	2·98	2·95	3·50	3·45
	3	1·71	1·71	2·07	2·07	2·51	2·51	2·83	2·83
	5	1·68	1·68	2·03	2·02	2·44	2·43	2·73	2·72
	10	1·66	1·66	1·99	1·99	2·38	2·38	2·65	2·65
10	1	1·81	1·81	2·25	2·24	2·83	2·81	3·29	3·24
	3	1·70	1·70	2·05	2·05	2·48	2·47	2·79	2·78
	5	1·68	1·68	2·01	2·01	2·41	2·41	2·70	2·69
	10	1·66	1·66	1·99	1·98	2·38	2·37	2·64	2·63

* Computed from the t-approximation suggested above.
† Computed by Lord, using numerical quadrature.

To test for the significance of (*a*) the difference between the mean of a sample, $\bar{x}$, and a pre-assigned value of a population mean, μ, (*b*) the difference between two sample means $\bar{x}_1$, $\bar{x}_2$, we can form test criteria using the mean range of the observations in the sample or samples concerned. It is known that the sample mean is statistically independent of the range in the sample. The mean $\bar{x}$ is, therefore, also independent of the mean range $\bar{w}$ of the subgroups of the samples.

---

* We have considered here the double-tailed test, rejecting the null hypothesis if $|t| \geqslant t_{\frac{1}{2}\alpha}$, where $P\{|t| \geqslant t_{\frac{1}{2}\alpha}\} = \alpha$.

Thus we can represent the distribution of $\bar{w}_{m,n}/(c\sigma)$ by that of $\chi/\sqrt{\nu}$. Since s/σ follows the distribution of $\chi/\sqrt{\nu}$, s being the sample root-mean-square estimate of σ based on ν degrees of freedom, it follows that $\bar{w}_{m,n}$ is distributed approximately as cs.

From the mean and standard deviation of range given in Pearson's Table A (1932), the values of M and V in (1) are known, and using these in (5), ν is obtained and then c from (6). We give here (Table 1) a short table of d_n, ν and c for $n = 3$ to 10, $m = 1$ to 5.

Table 1. *Values of d_n, ν and c for the χ approximation to $w_{m,n}$*

Size of sample n	d_n	No. of samples									
		$m = 1$		$m = 2$		$m = 3$		$m = 4$		$m = 5$	
		ν	c	ν	c	ν	c	ν	c	ν	c
3	1·6926	1 934	1·9164	3·850	1·8049	5·674	1·7684	7·499	1·7408	9·312	1·7385
4	2·0588	2·951	2·2374	5·705	2·1505	8·450	2·1204	11·191	2·1052	13·931	2·0960
5	2·3259	3·828	2·4812	7·474	2·4048	11·107	2·3788	14·732	2·3657	18·355	2·3578
6	2·5344	4·692	2·6721	9·169	2·6001	13·641	2·5812	18·109	2·5696	22·576	2·5626
7	2·7044	5·499	2·8295	10·779	2·7677	16·052	2·7468	21·324	2·7362	26·596	2·7299
8	2·8472	6·259	2·9630	12·297	2·9056	18·328	2·8850	24·358	2·8768	30·386	2·8707
9	2·9700	6·989	3·0778	13·753	3·0245	20·511	3·0064	27·270	2·9974	34·025	2·9920
10	3·0775	7·689	3·1789	15·151	3·1287	22·609	3·1117	30·066	3·1032	37·523	3·0981

N.B. $w_{m,n}/\sigma$ is distributed approximately as $c\chi/\sqrt{\nu}$, where c is a scale factor and ν a fractional degree of freedom.

Unfortunately, there is at present no direct means of judging the accuracy of the above approximation. When the exact moments of w_n above the second are computed, it will be possible to derive the higher moments of $\bar{w}_{m,n}$ and so compare them with the moments of our approximate s-distribution. The probability integral and percentage points of the distribution of $\bar{w}_{m,n}$ can be obtained approximately from those of the χ and hence from those of the χ^2-distribution with degrees of freedom given by the ν of equation (5).

Turning to Table 1, we note the way in which c tends to d_n from above as n increases; further, we see that the equivalent, 'fractional' degrees of freedom, ν, of the approximation are always rather less than $m(n-1)$, the figure appropriate were we to estimate σ from the within-group sums of squares.

3. DISTRIBUTION OF $x/\bar{w}$

Let x be a variate distributed normally with mean zero and standard deviation σ, and let s be an independent root-mean-square estimate of σ based on ν degrees of freedom. In § 2 it was seen that the mean range could be used for estimating σ. Therefore in parallel with the ratio t, we may take the ratio

$$u = \frac{x}{\bar{w}/d_n}, \tag{7}$$

where $\bar{w}$ stands for $\bar{w}_{m,n}$* and $d_n\sigma$ is the expectation of the range in samples of n.

Lord (1947, 1950) has considered this distribution and developed the 'u-test', giving several of its applications. A wide range of tables of percentage points have been computed by him by methods of quadrature and interpolation. Daly (1946) also deals with this test, but he takes a single range instead of the mean range.

---

* In the rest of this section $\bar{w}$ will be used for $\bar{w}_{m,n}$.

We will now employ the approximate distribution of $\bar{w}$ obtained in § 2 to derive an approximation to the distribution of u. Since $\bar{w}$ is represented by cs, we see that the distribution of u can be approximated by that of

$$u = \frac{d_n x}{cs} = \frac{d_n}{c} t, \tag{8}$$

where t is based on ν degrees of freedom, c and ν being given by equations (5) and (6). In other words, $u\{1 + 1/(4\nu) + 1/(32\nu^2) - 5/(128\nu^3)\}$ is distributed approximately as 'Student's' t with degrees of freedom ν given by equation (5). It is therefore possible to find approximate percentage points for u by interpolating for a fractional ν in any table of the percentage points of t. Using this method, we have calculated the percentage points of the distribution for significance levels* $\alpha = 0{\cdot}10$, $0{\cdot}05$, $0{\cdot}02$ and $0{\cdot}01$ for several values of n and m. They are shown in Table 2, and Lord's values are given alongside for comparison. The good agreement, even for low values of α and for $m = 1$, i.e. for a single range, suggests that the distribution of u is closely approximated by that of $d_n t/c$. We may also infer that the χ-approximation to the distribution of $\bar{w}$ is fairly good.

Table 2. *Values of the percentage points of the distribution of* $u = d_n x/\bar{w}_{m,n}$

Size of sample n	No. of samples m	$\alpha = 0{\cdot}10$		$\alpha = 0{\cdot}05$		$\alpha = 0{\cdot}02$		$\alpha = 0{\cdot}01$	
		Approx.*	Lord†	Approx.*	Lord†	Approx.*	Lord†	Approx.*	Lord†
5	1	2·02	2·02	2·64	2·63	3·58	3·56	4·39	4·38
	3	1·75	1·75	2·15	2·15	2·65	2·65	3·03	3·02
	5	1·71	1·71	2·07	2·07	2·51	2·51	2·83	2·83
	10	1·68	1·68	2·01	2·01	2·42	2·42	2·70	2·70
8	1	1·85	1·85	2·33	2·32	2·98	2·95	3·50	3·45
	3	1·71	1·71	2·07	2·07	2·51	2·51	2·83	2·83
	5	1·68	1·68	2·03	2·02	2·44	2·43	2·73	2·72
	10	1·66	1·66	1·99	1·99	2·38	2·38	2·65	2·65
10	1	1·81	1·81	2·25	2·24	2·83	2·81	3·29	3·24
	3	1·70	1·70	2·05	2·05	2·48	2·47	2·79	2·78
	5	1·68	1·68	2·01	2·01	2·41	2·41	2·70	2·69
	10	1·66	1·66	1·99	1·98	2·38	2·37	2·64	2·63

* Computed from the t-approximation suggested above.
† Computed by Lord, using numerical quadrature.

To test for the significance of (*a*) the difference between the mean of a sample, $\bar{x}$, and a pre-assigned value of a population mean, μ, (*b*) the difference between two sample means $\bar{x}_1$, $\bar{x}_2$, we can form test criteria using the mean range of the observations in the sample or samples concerned. It is known that the sample mean is statistically independent of the range in the sample. The mean $\bar{x}$ is, therefore, also independent of the mean range $\bar{w}$ of the subgroups of the samples.

---

* We have considered here the double-tailed test, rejecting the null hypothesis if $|t| \geqslant t_{1\alpha}$, where $P\{|t| \geqslant t_{1\alpha}\} = \alpha$.

In case (*a*), if the sample of N can be divided randomly into m subsamples of n observations each,* then we form the ratio

$$v = \frac{cu}{d_N} = \frac{c\sqrt{N}(\bar{x}-\mu)}{\bar{w}_{m,n}}, \tag{9}$$

which will be distributed as t in virtue of (8), with ν degrees of freedom. In case (*b*), if the two samples of sizes N_1, N_2 can be randomly divided into m_1, m_2 subsamples of size n,* then the mean of the $k = m_1 + m_2$ ranges is taken and the statistic

$$v = \frac{c(\bar{x}_1 - \bar{x}_2)}{\bar{w}_{k,n}\sqrt{(1/N_1 + 1/N_2)}} \tag{10}$$

follows the distribution of t with ν degrees of freedom. Thus in both cases we refer v to the t-probability scale. ν and c are given by equations (5) and (6), and in certain cases their values may be taken from Table 1. We shall illustrate the procedure by an example.

Example 1. Consider the data of Example 5 given by Lord (1947, p. 60). The mean strength (in lb.) of cotton yarn in two samples, A (with $N_1 = 30$) and B (with $N_2 = 20$), and the ranges of the five subsamples, of ten observations each, are:

	Sample A	Sample B
Mean	28·90	27·40
Range	3·0, 5·0, 5·0	4·0, 3·5

Hence, mean range, $\bar{w}_{5,10} = 4\cdot1$, difference of means $\bar{x}_1 - \bar{x}_2 = 1\cdot5$.

From Table 1, $c = 3\cdot098$, $\nu = 37\cdot52$. The computed value of v in (10) is 3·926. This falls beyond the 0·1 % point for t, which for $\nu = 37\cdot52$ is at about 3·6, and therefore the difference in strength of the two batches of yarn is certainly significant.

The value of 'Student's' t for the data of this example was found to be 3·7, which again exceeds the 0·1 % point of t for 48 degrees of freedom.

4. DISTRIBUTION OF $W/\bar{w}$

Suppose W_N is the range in a sample of N observations from a normal population with mean μ and standard deviation σ. Let there be m further independent samples of n drawn from the same population, and let the mean of their ranges be $\bar{w}_{m,n}$. Using the s-approximation to the distribution of $\bar{w}$ obtained in § 2, it follows that the ratio

$$g = \frac{cW_N}{\bar{w}_{m,n}} \tag{11}$$

will be distributed approximately as the ratio of (*a*) a range in a sample of N observations, to (*b*) an independent root-mean-square estimate of σ based on ν degrees of freedom, where ν and c are given by equations (5) and (6). This ratio is the so-called 'Studentized' range, $q = W_N/s$, the probability integral and percentage points of which have been tabled by Pearson & Hartley (1942).†

The accuracy of the above approximation to the range ratio $W/\bar{w}$ could only be judged when the exact percentage points are computed by methods similar to those used for $d_n x/\bar{w}$.

---

* In the preceding paper, Lord (1950) has shown that it may be worth while omitting a few observations in the estimation of σ, in order to ensure the breaking up of the remainder into appropriate equal groups.

† Note that the n of their tables corresponds to our N.

However, Hartley's lemma (1948) tells us that the maximum error in the probability integral of $W/\bar{w}$ will be less than the maximum error in the probability integral of $\bar{w}$, which may be fairly small judging from the agreement in the percentage points of $d_n x/\bar{w}$.

It may be noted that it would have been possible to employ the s-approximation for W_N as well as for $\bar{w}_{m,n}$ and so regard the square of the ratio $W/\bar{w}$ (when multiplied by the appropriate constant) as equivalent to a variance ratio, F. But as W_N is a single range, not a mean range, it seems probable that the 'Studentized' range approximation will be the more accurate.

5. APPLICATION OF THE DISTRIBUTION OF $W/\bar{w}$ TO THE ANALYSIS OF VARIANCE

Using the range-ratio, $W/\bar{w}$, we can formulate a test of significance, analogous to the F-test, applicable to analysis of variance. We shall consider a one-way classification where there are m groups with n observations in each.

Whether the systematic set-up* or the random set-up is relevant, if the null hypothesis is true, the observed group means are distributed normally with variance $= \sigma_0^2/n$, where σ_0^2 is the residual variance within the groups.

If W_m is the range of the group means and $\bar{w}_{m,n}$ the mean of the m group ranges, it can be shown that these two quantities are statistically independent. Hence, using our approximation for $\bar{w}_{m,n}$, the ratio $g = cW_m/\bar{w}_{m,n}$ is approximately distributed as W_m/s, where s is independent of W_m. Now W_m is 'Studentized' when it is divided by $s/\sqrt{n}$, the estimate of $\sigma_0/\sqrt{n}$. Hence $c\sqrt{n}\, W_m/\bar{w}_{m,n}$ is a 'Studentized' range of a sample of size m and with degrees of freedom ν.

Thus the 'range-test' for the hypothesis of no variability between groups consists in evaluating

$$q = \frac{W_m}{\bar{w}_{m,n}} c\sqrt{n}, \tag{12}$$

and testing whether or not it exceeds the 100α % point of the 'Studentized' range for m and ν. The procedure may best be illustrated by an example.

Example 2. The following table gives the yield of grain in bushels per acre, corresponding to seven varieties (Weatherburn, 1947, p. 213):

	V_1	V_2	V_3	V_4	V_5	V_6	V_7
	13	15	14	14	17	15	16
	11	11	10	10	15	9	12
	10	13	12	15	14	13	13
	16	18	13	17	19	14	15
	12	12	11	10	12	10	11
Means	12·4	13·8	12·0	13·2	15·4	12·2	13·4
Ranges	5	7	4	7	7	6	5

---

* The distinction between the systematic and random set-ups has been described by a number of authors; see, for example, Daniels (1939), Johnson (1948) and Patnaik (1949, p. 225).

We find $$W_7 = 3{\cdot}4, \quad \overline{w}_{7,5} = \tfrac{41}{7}.$$

Further $M = d_5 = 2{\cdot}32593$ and, from Pearson's Table A (1932, p. 416), the variance of w_5/σ is $(0{\cdot}8641)^2$. Hence, from equation (1), $V = (0{\cdot}8641)^2/7 = 0{\cdot}10667$. Introducing these values into equations (5) and (6), we find

$$\nu = 25{\cdot}61, \quad c = 2{\cdot}349.$$

Finally, from (12) we have $q = 3{\cdot}049$. This value is considerably below the upper 5 % limit for q given in Pearson & Hartley's tables, for $m = 7$, $\nu = 25{\cdot}61$, viz. $4{\cdot}53$. The data, therefore, do not suggest the existence of significant differences in the variety yields.

By the ordinary analysis of variance, the ratio of the mean square between varieties to that within varieties is found to be $1{\cdot}1$, while the $F_{0{\cdot}05}$ for $\nu_1 = 6$, $\nu_2 = 28$ is $2{\cdot}45$. Thus the range test leads to the same conclusion as the F-test.

6. THE POWER OF THE u, OR RANGE, TEST

The ratio u, of equation (7), may clearly be used as an alternative to t in tests of significance. Using our method of approximation, we shall now examine to what extent u is less efficient than t in detecting true differences in mean values. This may be done by deriving the power function of the u-test and comparing it with that of the t-test. In his second paper, Lord (1950) has made an exact comparison, using methods of quadrature; in what follows we shall apply a more approximate method, based on an adaptation of the non-central t-distribution.

In (7), x was assumed to have a mean, $\mu = 0$. If in fact $\mathscr{E}(x) = \mu \neq 0$, then u can be written in the form

$$u = \frac{(x-\mu)/\sigma + \mu/\sigma}{\overline{w}/(d_n\sigma)} = \frac{z+\mu/\sigma}{\overline{w}/(d_n\sigma)},$$

where $\mathscr{E}(z) = 0$ and $\operatorname{var}(z) = 1$. The ratio will follow what may be termed a non-central u-distribution. If now we use the cs approximation to $\overline{w}$ derived in § 2, it is seen that u may be written as

$$\frac{d_n}{c}\frac{z+\mu/\sigma}{s/\sigma},$$

or, expressed in another way, $$\frac{cx}{\overline{w}} = \frac{cu}{d_n} = \frac{z+\mu/\sigma}{s/\sigma}. \tag{13}$$

It follows from (13) that (a) if $\mu = 0$, $cx/\overline{w}$ is distributed as t with ν degrees of freedom; (b) if $\mu \neq 0$, it is distributed as non-central t with ν degrees of freedom and displacement μ/σ. c and ν will be given as before by equations (5) and (6). Thus the power function of Lord's u-test may be obtained, to the degree of approximation involved, from the power function of 'Student's' t-test, using the appropriate fractional degrees of freedom for ν.

Tables of the power of the t-test have been given by Neyman (1935). They are for the asymmetrical comparison, i.e. the case where we reject the null hypothesis if $t \geqslant t_\alpha$, where $P\{t \geqslant t_\alpha\} = \alpha$. Suppose that the estimate of variance is to be based on m groups of n observations; then using 'Student's' test, we shall have an estimate of σ^2 based on $m(n-1)$ degrees of freedom, while the equivalent degrees of freedom for Lord's test will be the fractional ν derived from equation (5) or, where possible, from our Table 1. Some comparisons of the two tests are made in Table 3 below, for displacements $\mu/\sigma = 1$, 2 and 3.

As we should expect, since $\nu < m(n-1)$ the power of the u-test to detect departures from the null hypothesis, $\mu = 0$, is always less than that of the t-test, but it is clear that, for the cases chosen, the differences in power are remarkably small.

Table 3. *Power functions of the u- and t-tests*

(Single-tailed test, with $\alpha = 0{\cdot}05$)

Size of sample n	No. of samples m	Values of μ/σ					
		1		2		3	
		u-test	t-test	u-test	t-test	u-test	t-test
5	3	0·240	0·242	0·592	0·595	0·879	0·881
	5	0·248	0·249	0·611	0·613	0·893	0·895
10	3	0·250	0·252	0·616	0·620	0·897	0·900
	5	0·253	0·254	0·623	0·624	0·901	0·902

It is necessary, of course, to consider what is the accuracy of the approximation to the non-central u-distribution. Since this paper was completed, Mr Lord has kindly put at my disposal his tables of computed values of the exact power function and the 'standardized error' of his test. His calculations are for the symmetrical or double-tailed test; here, we reject the null hypothesis if $|t| \geqslant t_{\frac{1}{2}\alpha}$, where $P\{|t| \geqslant t_{\frac{1}{2}\alpha}\} = \alpha$.

The following comparisons have been made:

(*a*) *Case* $m = 1$, $n = 10$

Table 1 above shows that, for the approximation, $\nu = 7{\cdot}689$. Interpolating with this fractional value in Lord's Table 1 of the power function of 'Student's' t (p. 67 above), we get the approximate values of the power shown below for $\mu/\sigma = 2$, 3 and 4. These are compared with the true values, taken from Lord's Table 3 (p. 70 above).

Power of double-tailed tests, with a significance level, $\alpha = 0{\cdot}05$

μ/σ	2	3	4
u-test: Approximation	0·417	0·743	0·934
True value	0·421	0·747	0·935
t-test: 9 degrees of freedom	0·431	0·761	0·944

In this case, with $m = 1$, the approximation was not expected to be very good. Let us make comparisons for $m > 1$. Here Lord gives only the standardized errors, that is to say, the values of $\rho = \mu/\sigma$ for which the power equals $1 - \alpha$. Thus $\rho_{0\cdot05}$ is the value of μ/σ for which there is a 0·95 chance of establishing significance, using the test at the 5 % level.

(*b*) *Case* $m = 2$, $n = 10$

Table 1 shows, for our approximation, that $\nu = 15{\cdot}151$. Using this value to interpolate in Lord's Table 2, it is found that $\rho_{0\cdot05} = 3{\cdot}86$. The true value given by Lord in his Table 5 is also 3·86.

(*c*) *Case* $m = 4$, $n = 5$

For the approximation, $\nu = 14{\cdot}732$ and hence from Lord's Table 2 (p. 68), $\rho_{0{\cdot}05} = 3{\cdot}87$, again agreeing to the second decimal place with the true value in his Table 5.

(*d*) *Case* $m = 5$, $n = 3$

For the approximation, $\nu = 9{\cdot}312$, giving $\rho_{0{\cdot}05} = 4{\cdot}05$, agreeing with Lord's computed value.

(*e*) *Case* $m = 2$, $n = 5$

For the approximation, $\nu = 7{\cdot}474$, giving $\rho_{0{\cdot}05} = 4{\cdot}19$, against Lord's value of $4{\cdot}18$.

There seems little doubt that, if a mean range can be used to estimate σ, the approximation will be very good.

7. THE POWER OF THE TEST BASED ON THE RATIO OF RANGES, $W/\bar{w}$

Turning now to the range-ratio test in the analysis of variance considered in §§4 and 5, we have to distinguish between the two set-ups—the random and the systematic—for considerations of power. Let us take the former and the simple case of m groups with n observations in each. If, as before, σ^2 is the variance between the groups and σ_0^2 the variance within the groups, the observed group means will have a variance of $\sigma^2 + \sigma_0^2/n$. With the usual notation, when $\sigma^2 \neq 0$, $g = cW/\bar{w}$ is approximately distributed as

$$\frac{W}{s} = \frac{\sqrt{\left(n\dfrac{\sigma^2}{\sigma_0^2} + 1\right)}}{\sqrt{n}} \frac{W}{s\sqrt{\left(\dfrac{\sigma^2}{\sigma_0^2} + \dfrac{1}{n}\right)}}$$

$$= \frac{\sqrt{\left(n\dfrac{\sigma^2}{\sigma_0^2} + 1\right)}}{\sqrt{n}}\, q,$$

where q is the 'Studentized' range of a sample of size m and with ν degrees of freedom. Hence the power of the g-test based on $W/\bar{w}$, to detect the presence of between-group variability measured in terms of σ^2/σ_0^2, can be obtained using the Pearson & Hartley (1942) table of the probability integral of q.

This result could be compared with the power of the analysis of variance F-test under the random set-up (see Johnson, 1948).

On the other hand, if the systematic set-up is valid, then the alternative hypothesis states that the m group means have different expectations, but have the same variance σ_0^2/n. Now, their range W is what may be termed a non-central range, and so $W/\bar{w}$ will be distributed as the ratio of a non-central range to the average of m central ranges. The distribution function of non-central range is found to be the sum of a number of multiple integrals, and except in the case of sample size of 2 and 3 is extremely complicated. The distribution of the ratio of non-central range to mean-central range is not easy to handle even with our approximations.

REFERENCES

DANIELS, H. E. (1939). *J. R. Statist. Soc. Suppl.* **6**, 186.
DALY, J. F. (1946). *Ann. Math. Statist.* **17**, 71.
GRUBBS, F. E. & WEAVER, L. C. (1947). *J. Amer. Statist. Ass.* **42**, 224.
HARTLEY, H. O. (1942). *Biometrika*, **32**, 334.
HARTLEY, H. O. (1948). *Biometrika*, **35**, 417.
JOHNSON, N. L. (1948). *Biometrika*, **35**, 80.
LORD, E. (1947). *Biometrika*, **34**, 41.
LORD, E. (1950). *Biometrika*, **37**, 64.
McKAY, A. T. & PEARSON, E. S. (1933). *Biometrika*, **25**, 415.
NEYMAN, J. (1935). *J. R. Statist. Soc. Suppl.* **2**, 131.
PATNAIK, P. B. (1949). *Biometrika*, **36**, 202.
PEARSON, E. S. (1926). *Biometrika*, **18**, 173.
PEARSON, E. S. (1932). *Biometrika*, **24**, 404.
PEARSON, E. S. (1943). *Biometrika*, **33**, 89.
PEARSON, E. S. & HARTLEY, H. O. (1942). *Biometrika*, **32**, 301.
TIPPETT, L. H. C. (1925). *Biometrika*, **17**, 364.
WEATHERBURN, C. E. (1947). *Mathematical Statistics*. Cambridge University Press.

[*Editorial Note.* Since this paper was accepted for publication, a paper has been published by D. R. Cox (*J. Roy. Statist. Soc.*, Series B, 1949, pp. 101–114), in which the distribution of range is approximated to by using a χ^2 (or Γ) distribution with the correct first two moments. This method is distinct from that of Patnaik, who uses the distribution of χ. When testing whether two variances are equal, Cox's approach leads to the ratio of two ranges or mean ranges being distributed as F, whereas Patnaik's approach leads to the use of the tables of percentage points of 'Studentized' range. Cox is mainly concerned with applications in problems of sequential analysis.]

distributions were all Pearson-type curves, and while these cannot represent all forms of departure from the normal that may occur in the tails of frequency distributions, they cover a wide variety of shapes; four of the six population curves are shown on p. 122 of Pearson (1931). It is clear that, at least, up to samples of size $n = 10$, the expectation of w_n/d_n, and therefore of $\bar{w}_{m,n}/d_n$, remains remarkably insensitive to changes in the population β_1 and β_2.* The value of σ_w, which, however, is less important than that of the expectation in its effect on the distribution of u, does change considerably from the normal-theory value but so also,

Table 1. *Expectations and standard errors of the range estimator of σ, namely, w_n/d_n*

(*a*) Symmetrical distributions (types II and VII)

β_2		$n = 5$	$n = 10$	$n = 20$
1·8 (Rectangle)	Expectation S.E.	*0·993* *0·265*	*0·921* *0·125*	*0·839* *0·058*
2·5 (II)	Expectation S.E.	1·008 ± 0·015 0·338 (500)	0·983 ± 0·010 0·232 (500)	0·965 ± 0·007 0·147 (500)
3·0 (Normal)	Expectation S.E.	*1·000* *0·371*	*1·000* *0·259*	*1·000* *0·195*
4·1 (VII)	Expectation S.E.	1·015 ± 0·019 0·434 (500)	1·002 ± 0·013 0·290 (500)	1·060 ± 0·012 0·275 (500)
7·1 (VII)	Expectation S.E.	0·965 ± 0·015 0·473 (1000)	0·986 ± 0·018 0·399 (500)	1·046 ± 0·016 0·348 (500)

(*b*) Skew distributions (types III and I)

β_1, β_2		$n = 5$	$n = 10$	$n = 20$
0, 3·0 (Normal)	Expectation S.E.	*1·000* *0·371*	*1·000* *0·259*	*1·000* *0·195*
0·2, 3·3 (III)	Expectation S.E.	1·009 ± 0·013 0·410 (1000)	0·999 ± 0·012 0·259 (500)	0·990 ± 0·009 0·205 (500)
0·5, 3·75 (III)	Expectation S.E.	0·985 ± 0·012 0·384 (1000)	0·967 ± 0·012 0·265 (500)	1·007 ± 0·007 0·222 (1000)
1·0, 3·8 (I)	Expectation S.E.	0·968 ± 0·012 0·392 (1000)	0·963 ± 0·013 0·296 (500)	0·933 ± 0·009 0·210 (500)

Note. The results are given in terms of the population standard deviation as unit. The figures in italics are theoretical, the others were derived from sampling experiments. The sampling standard errors of the expectations are calculated from $s_e/\sqrt{N}$, where s_e is the standard error of w_n/d_n observed in the N random samples. The value of N (500 or 1000) is shown in parentheses.

---

* The symmetrical distributions with $\beta_2 = 1{\cdot}8$ (the rectangle) and $\beta_2 = 7{\cdot}1$ and the skew distribution with $\beta_1 = 1{\cdot}0$, $\beta_2 = 3{\cdot}8$ are exceptional forms not commonly met in practice. Abruptness at one or both terminals will reduce the expectation of the range estimator when the sample size increases. This effect appears in the case of the rectangular distribution and also for the type I curve, which is very steep at the start.

of course, does the standard error of s. It is known that as n increases the standard error of s equals approximately the normal theory value multiplied by $\sqrt{\{(\beta_2-1)/2\}}$. When $n=20$, the same factor may be seen to give roughly the empirical standard errors of the range estimator.

8. In a recent paper, K. J. Shone (1949) has considered the relation between mean range and population standard deviation in samples of $n = 2$ to 5, from a number of non-normal populations. He found exactly, for $n = 5$, the following values for $\mathscr{E}(w/d_5\sigma) = \bar{w}/(d_5\sigma)$:

Population	β_1	β_2	$\bar{w}/(d_5\sigma)$
A	0	2·67	0·975
D	0	1·78	0·979
G	0	4·63	0·926
I	0·58	3·46	0·958
K	0·85	3·22	0·937

The population distributions each containing only seven discrete frequencies, spaced at equal intervals, were quite unlike those used in the sampling experiments referred to above, and it is probably for this reason that the values of $\bar{w}/(d_5\sigma)$ differ considerably more from unity than those given in Table 1. The latter were based on samplings from populations with a much more finely graded variable, and it is unlikely that any 'grouping up' could bring the two series into agreement.

EFFECT OF ROGUE OBSERVATIONS

9. If the sample is large, a single outlying observation will in general have a much greater influence on the range than on the root-mean-square estimator of σ, but it is not immediately clear that this will be the case in a small sample. An illustration of the position can be obtained as follows. Suppose that $x_1, x_2, \ldots, x_n$ represent a random sample of n observations, arranged in ascending order of magnitude, drawn from a normal population having zero mean and unit S.D. Suppose now that an 'error' term a gets for some reason added to x_n, to make it $X_n = x_n + a$.*

The range becomes $w = X_n - x_1$, and the expectation of the range estimator is

$$\mathscr{E}(w/d_n) = 1 + a/d_n, \tag{6}$$

where d_n is the tabled corrective factor. The mean-square estimator of σ^2 will be

$$s^2 = \left\{\sum_{i=1}^{n-1}(x_i-\bar{x})^2 + (X_n-\bar{x})^2\right\}\Big/(n-1),$$

where

$$n\bar{x} = \sum_{i=1}^{n}(x_i) + a.$$

It can be shown that

$$\begin{aligned}\mathscr{E}(s^2) &= 1 + a^2/n + 2a\mathscr{E}(x_n)/(n-1),\\ &= 1 + a^2/n + ad_n/(n-1),\end{aligned} \tag{7}$$

where $\mathscr{E}(x_n) = \frac{1}{2}d_n$ is the expected derivation of the extreme individual in a sample of n from the population mean, when $\sigma = 1$.

---

* This device gives a rapid means of examining the effect of an outlying observation, more reasonably, the error should be added at random to any one of the n observations and not to the extreme. However, this would complicate the calculation of $\mathscr{E}(w)$.

10. When $a = a_1 \neq 0$, the value of $\mathscr{E}(s)$ is not immediately obtainable, but it seems reasonable to assume that, approximately,

$$\frac{\mathscr{E}(s \mid a = a_1)}{\mathscr{E}(s \mid a = 0)} = \left\{\frac{\mathscr{E}(s^2 \mid a = a_1)}{\mathscr{E}(s^2 \mid a = 0)}\right\}^{\frac{1}{2}} = \left\{1 + \frac{a_1^2}{n} + \frac{a_1 d_n}{n-1}\right\}^{\frac{1}{2}}. \tag{8}$$

In Table 2, the expressions (8) and (6) are compared for $n = 5$, 10, 20, 30 and $a = 0{\cdot}5$, 1·0, 2·0, the population standard deviation being taken as unit.

Table 2. *Showing the effect on estimates of σ of the addition of an error term, $a\sigma$, to the extreme observation in a sample. (σ taken as unity)*

Sample size n	$a = 0{\cdot}5$		$a = 1{\cdot}0$		$a = 2{\cdot}0$		d_n
	$\sqrt{\mathscr{E}(s^2)}$	$\mathscr{E}(w/d_n)$	$\sqrt{\mathscr{E}(s^2)}$	$\mathscr{E}(w/d_n)$	$\sqrt{\mathscr{E}(s^2)}$	$\mathscr{E}(w/d_n)$	
5	1·16	1·21	1·33	1·43	1·72	1·86	2·3259
10	1·09	1·16	1·20	1·32	1·44	1·65	3·0775
20	1·05	1·13	1·12	1·27	1·20	1·54	3·7349
30	1·04	1·12	1·08	1·24	1·19	1·49	4·0856

It is seen that for a given error term, a, the expectation of the range estimator is more at fault than that of the root-mean-square estimator, and that, relatively, this bias increases the larger the sample and the larger the error a. If there is a single outlier only, and we break up the sample into groups when using the range estimator, the bias is very much reduced. Thus if thirty observations are broken into three random groups of $n = 10$, it will be seen that $\mathscr{E}(\bar{w}_{3,10}/d_{10})$ has the following values:

a	0·5	1·0	2·0
$\mathscr{E}(\bar{w}_{3,10}/d_{10})$	1·05	1·11	1·22
$\sqrt{\mathscr{E}(s^2)}$ $(n = 30)$	1·04	1·08	1·19

The bias in this way is only a little greater than that obtained from the root-mean-square estimator, applied to the whole thirty observations. A mean range based on six groups of five, would be less affected still.

11. In conclusion, these results may be regarded as supporting the case for using range, with the adjustment appropriate for a normal population, as an estimator of standard deviation provided that groups of not more than about ten observations are taken.

REFERENCES

Gayen, A. K. (1949). *Biometrika*, **36**, 353.
Geary, R. C. (1947). *Biometrika*, **34**, 209.
Ghurye, S. G. (1949). *Biometrika*, **36**, 426.
Lord, E. (1947). *Biometrika*, **34**, 41.
Pearson, E. S. (1929). *Biometrika*, **21**, 259.
Pearson, E. S. (1931). *Biometrika*, **23**, 114.
Pearson, E. S. & Adyanthaya, N. K. (1928). *Biometrika*, **20 A**, 356.
Shone, K. J. (1949). *J. R. Statist. Soc.* B, **11**, 85.

ON THE STATISTICAL INDEPENDENCE OF QUADRATIC FORMS IN NORMAL VARIATES

By A. C. AITKEN, D.Sc., F.R.S.

1. Introductory

The t-test, the variance-ratio test and certain other tests of significance are valid only on condition that the linear, quadratic or bilinear forms concerned are statistically independent. For quadratic forms, the variates being normal, uncorrelated and of unit variance, the central theorem is that of A. T. Craig (Craig, 1943; Hotelling, 1944), namely:

Let x be a vector having for elements n uncorrelated normal variates each of unit variance: then the necessary and sufficient condition for the quadratic forms $x'Ax$ and $x'Bx$ to be statistically independent is that $AB = 0$.

Some years earlier W. G. Cochran had obtained (Cochran, 1934, p. 181) the equivalent but less easily applied condition that

$$|I-\alpha A-\beta B| = |I-\alpha A|\,|I-\beta B| \tag{1}$$

identically in the real indeterminates α, β. In the present note we shall derive Craig's criterion from Cochran's condition, and we shall point out some corollaries.

2. Proof of Craig's theorem

The joint moment-generating function or Laplace transform of $x'Ax$ and $x'Bx$ is found at once (Cochran, 1934; Aitken, 1940) to be

$$G(\alpha,\beta) = (2\pi)^{-\frac{1}{2}n}\int_{-\infty}^{\infty} \exp\left(-\tfrac{1}{2}x'x+\tfrac{1}{2}\alpha x'Ax+\tfrac{1}{2}\beta x'Bx\right)dx = |I-\alpha A-\beta B|^{-\frac{1}{2}}, \tag{1}$$

provided α and β are in a real region such that $I-\alpha A-\beta B$ is positive definite.

The necessary and sufficient condition for independence, namely, that

$$G(\alpha,\beta) = G(\alpha,0)\,G(0,\beta),$$

then yields the condition quoted in § 1 (1). For convenience we shall write it in the homogeneous form

$$\lambda^n|\lambda I-\rho A-\sigma B| = |\lambda I-\rho A|\,|\lambda I-\sigma B|, \tag{2}$$

where ρ, σ are the arbitrary (real) indeterminates, in a suitable region. This shows that the latent roots of the matrix pencil $\rho A+\sigma B$ are the latent roots of ρA together with those of σB, with the useful corollary that the rank of $\rho A+\sigma B$ is the sum of the ranks of ρA and of σB. Also in the first place, and without loss of generality, we may suppose that an orthogonal transformation has been applied to the pencil, of such a kind that acting on A alone it produces a diagonal canonical form with the non-zero latent roots of A, let us say α_i ($i = 1, 2, 3, \ldots, r$) appearing in leading position in a submatrix A_1. It is permissible therefore to take A as given in this shape. Further, this being done, it is possible that the rows later than the rth in the transformed pencil may be linearly dependent among themselves. In such a case we could apply a further orthogonal transformation to reduce to nullity any redundant rows, and the corresponding columns, thus removing from the problem any redundant variates.

The problem is thus simplified to one in which, for real and non-zero ρ and σ, the pencil $\rho A + \sigma B$ is of full rank $n = r + s$, where r is the rank of A, and s is the rank of B. Since the latent roots of the pencil are known, there exists an orthogonal transformation to canonical form, let us say

$$Q'(\rho A + \sigma B) Q = \begin{bmatrix} \rho A_1 & \cdot \\ \cdot & \sigma B_2 \end{bmatrix}, \tag{3}$$

where B_2 must be diag $[\beta_1 \beta_2 \ldots \beta_s]$, the β_j being the non-zero latent roots of B. Let us partition B accordingly and examine the determinant of the pencil. We must have, identically in σ and ρ,

$$\begin{vmatrix} \rho A_1 + \sigma B_{11} & \sigma B_{12} \\ \sigma B'_{12} & \sigma B_{22} \end{vmatrix} = \begin{vmatrix} \rho A_1 & \cdot \\ \cdot & \sigma B_2 \end{vmatrix}. \tag{4}$$

The coefficient of $\rho^r \sigma^s$ on each side of this gives

$$|B_{22}| = |B_2| = \beta_1 \beta_2 \ldots \beta_s. \tag{5}$$

Now B is of rank s, and so is B_{22}. Thus the following congruent transformation, involving a submatrix K, exists:

$$\begin{bmatrix} I & K' \\ \cdot & I \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B'_{12} & B_{22} \end{bmatrix} \begin{bmatrix} I & \cdot \\ K & I \end{bmatrix} = \begin{bmatrix} \cdot & \cdot \\ \cdot & B_{22} \end{bmatrix}. \tag{6}$$

Applying this to $B - \lambda I$ and constructing the characteristic polynomial of B, we have therefore

$$|B - \lambda I| = \begin{vmatrix} -\lambda(I + K'K) & -\lambda K' \\ -\lambda K & B_{22} - \lambda I \end{vmatrix}. \tag{7}$$

The determinant on the right of (7) has a factor λ^r; this merely illustrates the fact that r of the latent roots of B are equal to zero. The other factor of the determinant has $|I + K'K|\,|B_{22}|$ for its constant term, and this by matrix theory must be equal to $\beta_1 \beta_2 \ldots \beta_s$, that is, by (5), to $|B_{22}|$. Hence $|I + K'K| = 1$. Since $K'K$ is non-negative definite, this can hold only if $K = 0$. Hence finally by (6),

$$B = \begin{bmatrix} \cdot & \cdot \\ \cdot & B_{22} \end{bmatrix}. \tag{8}$$

Thus the submatrices A_1, the non-zero part of A, and B_{22}, the non-zero part of B, are *disjunct*; and so $AB = 0$, as was to be proved. The sufficiency of the condition now follows at once from (2).

3. EXTENSION TO CASE OF CORRELATED VARIATES

Next, let the elements of the vector x be variates obeying a normal law characterized by $x'V^{-1}x$, where V is the variance matrix. Since V is positive definite, it admits a real square root $V^{\frac{1}{2}}$.

The transformation $y = V^{-\frac{1}{2}}x$ then produces uncorrelated variates y, all of unit variance. The quadratic forms $x'Ax$ and $x'Bx$ become, in the new variates, $y'V^{\frac{1}{2}}AV^{\frac{1}{2}}y$ and $y'V^{\frac{1}{2}}BV^{\frac{1}{2}}y$. The necessary and sufficient condition for their independence is

$$V^{\frac{1}{2}}AV^{\frac{1}{2}} . V^{\frac{1}{2}}BV^{\frac{1}{2}} = 0, \quad \text{that is,} \quad AVB = 0,$$

the appropriate extension of Craig's theorem.

4. THE DEPENDENCE OF A QUADRATIC AND A LINEAR FORM

In the case of a linear form $a'x$ with zero mean and a quadratic form $x'Bx$, the author derived a condition (Aitken, 1940) of independence, namely, that

$$a'(V-\beta B)^{-1}a \quad \text{should be independent of } \beta. \tag{1}$$

This result, however, like that of § 1 (1), could have been pressed further. Let us square the linear form, obtaining the quadratic form $x'aa'x$. By § 3, the necessary and sufficient condition for this to be independent of $x'Bx$ is that

$$BVaa' = 0, \quad \text{which further yields} \quad BVaa'a = 0. \tag{2}$$

Cancelling the non-zero scalar $a'a$, we derive the simpler condition $BVa = 0$, which is easily seen to be sufficient as well as necessary.

5. THE INDEPENDENCE OF BILINEAR FORMS

Bilinear forms $y'Ax$, $y'Bx$ may be constructed from sets of normal variates either independent or intercorrelated. In either case we may pool the variates and then regard any bilinear form as a quadratic form in the combined set. Craig's theorem can then be applied, and leads to various subsidiary theorems, which can be enunciated if desired; in practice, however, it seems easier to apply the theorem in its primitive form and not to particularize.

We consider, however, a possible case. Let the elements of x be r variates, those of y be the complementary s variates in a set of $n = r+s$ variates having V for their variance matrix. Consider bilinear forms $y'Ax$ and $y'Bx$. Apart from a factor 2 they may be written as matrix products, quadratic forms, thus·

$$[x' \vdots y']\begin{bmatrix} \cdot & A' \\ A & \cdot \end{bmatrix}\begin{bmatrix} x \\ \cdots \\ y \end{bmatrix}, \qquad [x' \vdots y']\begin{bmatrix} \cdot & B' \\ B & \cdot \end{bmatrix}\begin{bmatrix} x \\ \cdots \\ y \end{bmatrix}. \tag{1}$$

Partitioning V suitably, we may write down Craig's condition of independence as

$$\begin{bmatrix} \cdot & A' \\ A & \cdot \end{bmatrix}\begin{bmatrix} V_{11} & V_{12} \\ V'_{12} & V_{22} \end{bmatrix}\begin{bmatrix} \cdot & B' \\ B & \cdot \end{bmatrix} = \begin{bmatrix} A'V_{22}B & A'V'_{12}B' \\ AV_{12}B & AV_{11}B' \end{bmatrix} = 0. \tag{2}$$

This yields an easily applied criterion, which simplifies further in the case when $V_{12} = 0$, that is, when the two sets of variates are correlated within, but not between, their respective variates.

6. COCHRAN'S THEOREM ON INDEPENDENT QUADRATIC FORMS

Thus far we have not specialized the quadratic forms. For example, we have not insisted that they should be non-negative definite. But in the theory of analysis of variance there is a theorem on quadratic forms (Cochran, 1934), in which conditions of definiteness are imposed. Formulated as a theorem on matrices, it is as follows:

Given (i) *that* $A_1+A_2+\ldots+A_k = I$, *where all the matrices* A_i *are real symmetric matrices of order* $n \times n$; (ii) *that the ranks* n_i *of the* A_i *sum to* n: *then the separate* A_i *are idempotent, and there exists an orthogonal matrix* H *such that the non-zero parts of* $H'A_iH$ *are disjunct diagonal submatrices, with unit diagonal elements.*

The theorem admits a proof on matrix lines. Let us first apply an orthogonal transformation H_1 which would reduce A_1 to leading diagonal canonical form. We then have

$$H'_1A_1H_1 = \begin{bmatrix} D_1 & \cdot \\ \cdot & \cdot \end{bmatrix}, \quad H'_1(A_2+A_3+\ldots+A_k)H_1 = \begin{bmatrix} I-D_1 & \cdot \\ \cdot & I \end{bmatrix}. \tag{1}$$

But, since the rank of a sum of matrices cannot exceed the sum of the ranks of the separate matrices, the rank of $A_2+A_3+\ldots+A_k$ cannot exceed $n_2+n_3+\ldots+n_k$; and since the rank of the lower submatrix I on the right of (1) is certainly $n_2+n_3+\ldots+n_k$, we must have $I-D_1=0$, that is, $D_1=I$. In a similar manner we can proceed, transforming each successive A_i, independently of its predecessors, to its disjunct canonical form of n_i unit elements. It follows that the A_i are idempotent, and that the quadratic forms $x'A_ix$ are independent sums of squares, are in fact functions of χ^2 type with respective degrees of freedom n_i.

7. THE INDEPENDENCE OF NON-NEGATIVE QUADRATIC FORMS

In a paper that has lately appeared (Matérn, 1949) it is proved that for *non-negative* quadratic forms $x'Ax$ and $x'Bx$ in normal variates a sufficient condition for independence is that the forms shall be *uncorrelated*. This, in the case of independent normal variates x of unit variance, is shown to be equivalent to the condition

$$\operatorname{tr} AB = 0, \tag{1}$$

where 'tr' denotes trace or sum of diagonal elements. If the variance matrix of the vector x is V, the appropriate extension of (1) is

$$\operatorname{tr} AVBV = 0. \tag{2}$$

Matérn's result, for the case considered, implies Craig's theorem. For in the first place since

$$\operatorname{tr} AVBV = \operatorname{tr}(V^{\frac{1}{2}}AV^{\frac{1}{2}}.\,V^{\frac{1}{2}}BV^{\frac{1}{2}})$$

we can proceed without loss of generality (§ 3) as if V were I. But then also

$$\operatorname{tr} AB = \operatorname{tr}(B^{\frac{1}{2}}A^{\frac{1}{2}}.A^{\frac{1}{2}}B^{\frac{1}{2}}),$$

in virtue of the fact that a non-negative definite symmetric matrix has a non-negative definite square root. Since the matrix on the right is non-negative definite, the vanishing of its trace implies the vanishing of *all its diagonal elements*, and this implies the vanishing of *all* the elements of $A^{\frac{1}{2}}B^{\frac{1}{2}}$. Hence $AB = A^{\frac{1}{2}}.A^{\frac{1}{2}}B^{\frac{1}{2}}.B^{\frac{1}{2}} = 0$.

Thus Matérn's criterion, in the non-negative definite case, is equivalent to Craig's, but simpler.

The outcome is a criterion for the independence of non-negative definite quadratic forms which is perhaps the simplest of all to apply, namely:

The necessary and sufficient condition for the independence of $x'Ax$, $x'Bx$ is that all the diagonal elements of AVB (or BVA) shall be zero.

REFERENCES

AITKEN, A. C. (1940). On the independence of linear and quadratic forms in samples of normally distributed variates. *Proc. Roy. Soc. Edinb.* **60**, 40–6.

COCHRAN, W. G. (1934). The distribution of quadratic forms in a normal system, with applications to the analysis of covariance. *Proc. Camb. Phil. Soc.* **30**, 178–91.

CRAIG, A. T. (1943). Note on the independence of certain quadratic forms. *Ann. Math. Statist.* **14**, 195–7.

HOTELLING, H. (1944). On a matric theorem of A. T. Craig. *Ann. Math. Statist.* **15**, 427–9.

MATÉRN, B. (1949). Independence of non-negative quadratic forms in normally correlated variables. *Ann. Math. Statist.* **20**, 119–20.

TWO COMBINATORIAL TESTS OF WHETHER A SAMPLE HAS COME FROM A GIVEN POPULATION

By F. N. DAVID

1. The question as to whether a sample has come from a given population may be answered in two different ways. We may take the sample and carry out tests of significance regarding the sample mean, the sample variance and so on, each time obtaining an answer which tells us about that particular facet of the composition of the population which is revealed by the given sample test criterion, but we obtain no picture of the population as a whole. The alternative method of approach is to specify the possible functional form of the population and the necessary parameters of this functional form, and by means of the χ^2 criterion to test whether the sample could have been drawn from such a population. This latter test, in which the alternatives to the hypothesis tested may be whatever one likes—different functional form and/or different values for the parameters—appears to the writer most truly to answer the question 'Could the sample have come from a given population?' The χ^2 test does, however, suffer from the defect that it can only be applied to large samples because of the approximations involved in the derivation of its sampling distribution. Two tests are proposed here. The first is an omnibus test in which all alternatives to the hypothesis tested are regarded as admissible and which may be used instead of χ^2 when the size of sample is very small. The second test, a variant of the first, is more sensitive because it allows us to specify broadly which departures from the hypothesis under test it is most important not to overlook.

2. It will be assumed that the hypothesis tested, which we shall denote by H_0, is that the population, from which a sample of N observations has been independently and randomly drawn, is a continuous distribution of specified functional form and with given parameters. We suppose that this population is divided into k groups or strata of equal probability, and we plot the sample points noting into which strata they fall. It is clear that if these points are very regularly dispersed among the k strata then we have a situation analogous to that producing a very small value for χ^2. For example, if we make $k = N$ and there is one sample point in each group, then expectation is exactly realized. On the other hand, if all the sample points fall into one or two groups there is obviously a big departure from expectation corresponding to a large value of χ^2. We shall take therefore the number of groups containing no sample units—or if it is preferred, the number of groups containing at least one sample unit—as our test criterion.

3. The mathematical analysis required to find the distribution of the zero, or the non-zero, groups is very simple. It is given in great generality by Laplace in his *Théorie des Probabilités* (see also F. N. David, 1949). A simplification of the function derived by Laplace was discussed and tabulated by E. M. Elderton & K. Pearson (1925) and again later by W. L. Stevens (1937). Stevens was concerned with a test for randomness, which, viewed from one aspect, is what the test proposed in § 2 is also. We postulate a box divided into k equal and identical compartments. N identical balls ($N \leqslant k$) are dropped randomly into the box, any number of balls being allowed to fall into any one compartment. The number of ways in

which these N balls can arrange themselves in any $k-t$ compartments, leaving t empty, can be shown to be equal to

$$^kC_t\Delta^{k-t}0^N,$$

where $\Delta^{k-t}0^N$ is the $(k-t)$th leading difference of the Nth power of the natural numbers $0, 1, 2, 3, \ldots$, etc. Hence the probability of t empty compartments, given N balls and k compartments, is

$$P\{t \mid k, N\} = \frac{^kC_t}{k^N}\Delta^{k-t}0^N.$$

This result can be shown to be true also for $N > k$.

4. The analogy between the box, compartments and balls, and the probability integral, the groups of equal area and the sample units is obvious. N, the number of balls, is fixed by the sample size, but it is not immediate what value to choose for k, i.e. into how many equal parts we should divide our probability integral. Since this proposed test is to be used in a way analogous to χ^2, we recall the two different types of hypothesis alternate to that tested by the χ^2 criterion. We test the hypothesis that the functional form of the population and/or the given values of the parameters are what they are postulated to be. If we use the upper tail of the χ^2 distribution as a critical region then we have in mind as an alternative that the functional form and/or the given values of the parameters may be different from those tested. If we use the lower tail as a critical region, then the alternative hypothesis is complex and is concerned with the randomness or independence of the sample used for the test. The same situations obtain for the zeroes test and it is necessary to bear them in mind in choosing the value of k.

If k is chosen to be less than N we curtail the range of the probability distribution of the zeroes without achieving anything in compensation. The test will be insensitive in both directions and zero groups will be few, even when the functional form of the population is the correct one. This would suggest making the number of groups large. However, if we take $k > N$ the test will become insensitive to alternatives of the kind picked out by a very low χ^2. For if $k > N$ there will be always at least $k-N$ empty groups, no matter how the balls are distributed. It would appear, therefore, that the best balance to the test will be obtained when the number of compartments is equal, or nearly equal, to the number of balls. We illustrate this point in Table 2 with a sample of 5. Clearly if $k = 5$ there would be the possibility of establishing significance at either tail, even when N is as small as 5. This

Table 2. *Variation of probabilities of zero groups as k, the number of compartments, varies. (Sample size $N = 5$)*

No. of zero groups	Compartments		
	3	5	8
0	0·6173	0·0384	
1	0·3704	0·3840	
2	0·0123	0·4800	
3		0·0960	0·2051
4		0·0016	0·5127
5			0·2563
6			0·0256
7			0·0002

is not the case for $N = 3$ or 8. This question of the appropriate number of compartments to use is, theoretically, capable of solution by the use of power function technique, and the attempt was made to do this using the power function given in a later section. It is, however, proposed to use the test only for small samples, where the χ^2 test is not valid, and with these small samples the effect of discontinuities renders accurate comparisons between the power of the test for fixed sample size, but varying numbers of compartments, almost impossible. Calculations were carried out, and the suggestion which comes from them is that, for a fixed sample size, increasing the number of compartments tends to increase the power of the test using the upper tail of the sampling distribution but decreases the power of the test using the lower tail. We cannot state that to take the number of compartments equal to the number of balls is the optimum solution to the problem—it is possible that there will be a different optimum solution for each sample size. We can, however, state that when $k = N$ the two tails of the distribution are reasonably well balanced and we shall adopt this for our test. The proposed test will consist therefore of the following steps:

(i) Divide the probability integral into as many equal groups as there are units in the sample. (This is easy for any well-tabulated function.)

(ii) Count the number of groups containing no sample units.

(iii) Refer to Table 1*b* to judge whether the number of zero groups is too small or too large. The function

$$P\{t \mid N\} = \frac{{}^N C_t \Delta^{N-t} 0^N}{N^N}$$

is tabulated in Table 1*a* and the (nominal) significance levels in Table 1*b* for $N = 3$ to 20.

5. The zeroes test presented above will undoubtedly be insensitive, just as χ^2 is insensitive, when compared with any test which is specifically designed to test a possible given departure from H_0. The power function for the case when H_0 is false and an alternative hypothesis H_1 is true may easily be written down. We shall denote by

$$P(a_1 a_2 \ldots a_j) = \frac{N!}{(1!)^{a_1} a_1! \, (2!)^{a_2} a_2! \ldots}, \quad \sum_{i=1}^{j} a_i = t, \quad \sum_{i=1}^{j} i a_i = N,$$

the elementary partition function giving the number of ways in which N balls may arrange themselves in t compartments so that a_1 compartments contain one ball each, a_2 compartments contain 2 balls each, ..., a_j compartments contain j balls each. This is the same function as that denoted by David & Kendall (1949) as the augmented symmetric function, and we note that

$$\Sigma P(a_1 a_2 \ldots a_j) = \frac{\Delta^t 0^N}{t!},$$

the sum being taken over all possible partitions of N by t. Let $p_1, p_2, \ldots, p_N$ be the probabilities of a sample unit falling into the first, second, ..., groups under the alternative hypothesis H_1. Writing P_t for the probability of t zero classes under H_1, we have

$$P_{N-1} = \Sigma p_i^N,$$
$$P_{N-2} = \sum_{r,s} P(a_r a_s) \sum_{i \neq j} \Sigma\, p_i^r p_j^s,$$
$$P_{N-3} = \sum_{r,s,m} P(a_r a_s a_m) \Sigma \sum_{i \neq j \neq h} \Sigma\, p_i^r p_j^s p_h^m,$$

....

$$P_0 = \sum_{i \neq j \neq} \Sigma \ldots \sum_{\neq f} p_i p_j \ldots p_f \quad (N \text{ summations}),$$

where obviously

$$\sum_{i=1}^{k} \alpha_i = 0,$$

and we write

$$\sum_{i=1}^{k} \alpha_i^2 = \alpha.$$

We may deduce then that

$$\mathscr{E}(z) = k\left(1-\frac{1}{k}\right)^N + \frac{N(N-1)}{2}\left(1-\frac{1}{k}\right)^{N-2}\alpha,$$

$$\mathscr{E}(z-\mathscr{E}(z))^2 = k\left[(k-1)\left(1-\frac{2}{k}\right)^N + \left(1-\frac{1}{k}\right)^N - k\left(1-\frac{1}{k}\right)^{2N}\right]$$
$$+N(N-1)\alpha\left[(k-2)\left(1-\frac{2}{k}\right)^{N-2} + \frac{1}{2}\left(1-\frac{1}{k}\right)^{N-2} - k\left(1-\frac{1}{k}\right)^{2N-2}\right],$$

if we neglect terms involving sums of powers of α_i^s for $s > 2$. We do not know what the distribution of z under H_1 will be, but for small departures from H_0 we shall not expect normality to be greatly upset. We shall therefore assume that the distribution of z under H_1 is normal with mean and variance as given immediately above.

To get the power of the χ^2 test we follow Patnaik (1949). If we assume that the probabilities under H_0 are $\pi_1, \pi_2, \ldots, \pi_k$ and under H_1 are $p_1, p_2, \ldots, p_k$, then Patnaik showed that the quantity

$$\sum_{\text{All groups}} \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$$

is distributed approximately as non-central χ^2 with parameter

$$\lambda = N\left(\sum_{i=1}^{k} \frac{p_i^2}{\pi_i} - 1\right)$$

and degrees of freedom $k-1$. He gave tables from which, knowing λ and $k-1$ and the significance level of rejection under H_0, the power of the χ^2 test to detect the alternative hypothesis H_1 can be evaluated. Similar tables have also been calculated by E. Fix (1949).

We shall take the case of a sample of 30, and for illustrative purposes we take the hypothesis, H_0, that the sample has come from a normal population with zero mean and unit standard deviation. H_1 will be that the sample has come from a normal population with zero mean and with standard deviation $\frac{4}{3}$. For the zeroes test we take $k = 25$ and we find for a 5 % level of rejection under H_0, that the power of the test to detect the alternative hypothesis is 0·183. For the χ^2 test we divide the unit normal curve into six groups of equal area, thus ensuring an expectation of 5 in each group. $\pi_i = \frac{1}{6}$ for $i = 1, 2, \ldots, 6$. The quantities p_i are calculated corresponding to the dichotomy under H_0. We obtain $\lambda = 2{\cdot}477$, $k-1 = 5$ and interpolation into Patnaik's tables gives us the power as approximately 0·189. We see then that the power under the two tests is comparable.

Both the approximations to the power functions given above will hold (strictly) when the alternative hypothesis is not very different from the hypothesis tested. Numerical calculation has shown that Patnaik's approximation to the power function of χ^2 will give reasonable results for the power even when the alternate hypothesis is some way removed from that tested and in consequence the power is large. This unfortunately will not be the case for the approximation proposed for the power of the zeroes test and another approach is necessary. As the power of the test increases it will be found that a few of the p_i's become large and the remainder very small. For example, when the hypothesis alternate to that tested is that there has been a shift in the population mean to the right, then as the divergence becomes large $p_k \to 1$ and all the other p's tend to 0. This suggests that we expand the moments of z in ascending powers of the p's, most of which will quickly become negligible. Thus

$$\mathscr{E}(z) = \sum_{i=1}^{k} (1-p_i)^N = k - {}^NC_1 \Sigma p_i + {}^NC_2 \Sigma p_i^2 - {}^NC_3 \Sigma p_i^3 + \ldots,$$

$$\mathscr{E}(z(z-1)) = \sum_{\substack{i=1\\i\neq j}}^{k} \sum_{j=1}^{k} (1-p_i-p_j)^N = k(k-1) - {}^NC_1 \sum_{i\neq j}\Sigma (p_i+p_j) + {}^NC_2 \sum_{i\neq j}\Sigma (p_i+p_j)^2 - \ldots.$$

The double sums may be reduced to single sums and we have

$$\begin{aligned}
\sum_{i\neq j}\Sigma (p_i+p_j) &= k-1+\sum_i (k-1)\,p_i,\\
\sum_{i\neq j}\Sigma (p_i+p_j)^2 &= 2+2\sum_i (k-2)\,p_i^2,\\
\sum_{i\neq j}\Sigma (p_i+p_j)^3 &= 6\sum_i p_i^2 + \sum_i (2k-8)\,p_i^3,\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots
\end{aligned}$$

by making use of tables of symmetric functions. This procedure will be adequate for large divergencies from the hypothesis tested. It should, however, be used with caution because the assumption that z is normally distributed under H_1 need not necessarily be true in these cases.

9. The test for the counting of the number of zero classes supposes that the functional form of the population under H_0 is specified and the parameters also. But if sample moments are used as estimates of the parameters, the distribution of the number of zero classes will no longer be that given above owing to the restriction of the sample space. The test is therefore limited in application, a fault which it shares with most goodness of fit tests which have been proposed, the classic exception of course being χ^2. Recent work by David & Johnson (1948) on the probability integral transformation suggests that when x is subjected to two sample moment restraints, the transformed variable y defined in § 6 cannot be regarded as being effectively rectangularly distributed until $N \geqslant 30$. The procedure suggested above therefore will probably not lead to grave risk of error when one or two moment restraints are placed on the sample space, provided the sample size is greater than 30. Whether one would wish to use the zeroes test in such a case is a matter of opinion; probably the simplest thing would be to apply the χ^2 test as suggested earlier.

10. It will be recognized that the zeroes test proposed above is another addition to those tests of significance and of goodness of fit which are based on the probability integral transformation. E. S. Pearson (1938) discussed the theory of tests based on this transformation in some detail and pointed out that in order to overcome the lack of power associated with 'omnibus' tests it is necessary to define the way in which there may be departures from rectangularity, in order to devise tests which will be sensitive to these departures. We shall still consider, in this second test proposed below, the configuration of the zero classes in the sample, but we shall associate certain broad types of departure from the rectangular hypothesis with certain configurations of the zero classes, thus going some way towards satisfying the requirements laid down by E. S. Pearson. Suppose, for the sake of example, it is desired to use the zeroes test to take account of a possible shift in location of the population mean. If all the compartments to the right of the distribution are filled and none of those on the left, then it would be reasonable to suspect a shift in the population mean to the right. Again, if all the compartments in the centre of the distribution are filled, and none of those towards the two tails we might suspect a reduction in the population standard deviation, and so on.

We shall therefore base out second test on the configuration of the zero compartments in k specified compartments. These k will be selected before the analysis of the data, as being the most likely for an alteration in the number of zeroes to occur in them if the hypothesis tested is not true. We have again to face the question as to what should be the value of k to aim at. Presumably, arguing as before, we should now aim at making k, the number of specified compartments, equal to N, the number in the sample, and it will be undesirable to have the number of unspecified compartments, $>N$ or $<N$. We shall therefore divide the probability integral into $2N$ compartments and we shall choose $k = N$ of these for the test. In this test the presence of too many zero classes in the specified N groups will lead to rejection of the hypothesis tested. The presence of too few zero classes will not lead, however, to a suspicion of a lack of randomness, as it did in the test proposed earlier, but probably to a questioning as to whether the alternative hypothesis proposed is in fact the appropriate one to take into consideration.

Thus, if a change in the mean is suspected, we shall consider the configuration of the zeroes in the left-hand N or the right-hand N compartments, according to the direction in which the change is suspected. If we wish to take account of a change in the standard deviation, the mean being the same, then we shall divide the $2N$ compartments into three classes, the sum of the compartments in the outer two classes being N, and so on.

11. Our problem then corresponds to that of dropping N balls at random into $2N$ compartments and of finding the probability of $0, 1, 2, \ldots, N$ zeroes in any specified N compartments. It is clear that the total number of arrangements of the N balls in the $2N$ compartments will be $(2N)^N$. Let z be the number of zeroes in the specified N compartments and let A_z be the number of ways in which z may arise. We have then

$$A_N = \frac{N!}{(N-1)!}\frac{\Delta O^N}{1!} + \frac{N!}{(N-2)!}\frac{\Delta^2 O^N}{2!} + \ldots = \sum_{r=1}^{N} \frac{N!}{(N-r)!}\frac{\Delta^r O^N}{r!} = N^N,$$

$$A_{N-1} = \sum_{r=1}^{N} rN \frac{N!}{(N-r+1)!}\frac{\Delta^r O^N}{r!} = {}^N C_1 \Delta N^N = {}^N C_1[(N+1)^N - N^N],$$

and generally

$$A_{N-s} = \frac{N!}{(N-s)!}\sum_{r=s}^{N} {}^r C_s \frac{\Delta^r O^N}{r!}\frac{N!}{(N-r+s)!} = {}^N C_s \Delta^s N^N,$$

with $\quad A_0 = N!.$

The probabilities of $0, 1, 2, \ldots, N$ zeroes in the given N compartments are given by dividing the appropriate terms by $(2N)^N$. The probability function is tabulated in Table 4 *a* and *b*.

12. *Example I.* A sample giving values -8, $+7$, $+1$, $+10$, $+1$ is assumed to have been drawn from a normal population with mean, $\xi = 0$, and with standard deviation $\sigma = 10$. Test this hypothesis against alternatives which assume that $\xi > 0$.

From Table I of *Tables for Statisticians and Biometricians*, Part I, we read off the lines of dichotomy to be $-\infty$, $-12{\cdot}8$, $-8{\cdot}4$, $-5{\cdot}2$, $-2{\cdot}5$, 0, $2{\cdot}5$, $5{\cdot}2$, $8{\cdot}4$, $12{\cdot}8$, $+\infty$, and find that sample points fall in groups 3, 6, 8 and 9 counting from left to right. For alternative hypotheses making $\xi > 0$ we take the left-hand five compartments, in which there are four zeroes, and find from Table 4*a* below that this is not an uncommon configuration. In fact, the probability of four or more than four zeroes is 0·264. Table 4*b* shows that for a sample of size 5 significance will only be established if all five compartments specified are empty.

Example II. A sample of 4 units is assumed to have been drawn from a normal population with mean $\xi = 0$ and standard deviation $\sigma = 10$. The hypotheses alternative to that tested assume that there has been a reduction in the population standard deviation. Our lines of dichotomy are $-\infty$, $-11{\cdot}5$, $-6{\cdot}7$, $-3{\cdot}2$, 0, $+3{\cdot}2$, $+6{\cdot}7$, $+11{\cdot}5$, $+\infty$ and we decide to take groups 1, 2, 7 and 8 to test our hypothesis. The sample values are $-6, 0, 2, 2$.

We choose our test compartments *before* looking at the sample values, an important point. It is seen that no sample points fall into the test groups so that $z = 4$ and we may therefore reject the hypothesis that the parent population is described by a normal curve with mean $\xi = 0$ and standard deviation $\sigma = 10$ and proceed under the assumption that the normal curve describing the population will have $\xi = 0$ and $\sigma < 10$. In proceeding in this way, we shall in fact be wrong on the average six times in every 100 occasions.

13. The factorial moments of the distribution of z may be derived in straightforward fashion. We know that

$$P_{N-s} = \frac{N!}{(N-s)!\,s!}\frac{\Delta^s N^N}{(2N)^N} \quad \text{for} \quad s = 0, 1, \ldots, N.$$

Writing
$$(N-s)^{[k]} = (N-s)(N-s-1)\ldots(N-s-k+1),$$
it is immediate that
$$\begin{aligned}\mu'_{[k]} &= \sum_{s=0}^{N} (N-s)^{[k]} P_{N-s} = \frac{N^{[k]}}{(2N)^N}\sum_{s=0}^{N-k} {}^{N-k}C_s \Delta^s N^N \\ &= \frac{N^{[k]}}{(2N)^N}(1+\Delta)^{N-k} N^N = \frac{N^{[k]}}{(2N)^N} E^{N-k} N^N \\ &= N^{[k]}\left(1-\frac{k}{2N}\right)^N.\end{aligned}$$

We notice that when $k = 0$ the result is unity as expected. When $k = 1$ we have

$$\mu'_1 = N\left(1-\frac{1}{2N}\right)^N,$$

and we deduce from $k = 2$ that

$$\mu_2 = N^2\left[\left(1-\frac{1}{N}\right)^N - \left(1-\frac{1}{2N}\right)^{2N}\right] - N\left[\left(1-\frac{1}{N}\right)^N - \left(1-\frac{1}{2N}\right)^N\right],$$

moments which bear a certain similarity to those quoted for the first test. Higher moments can be deduced, if required, from the general formula given above.

14. The shape of the distribution of z for $N = 10$ (given in Table 4*a*) suggests that the normal curve may be a good approximation. A comparison between these exact values and the values obtained from a normal curve with moments as given in the preceding section is given in Table 5. A continuity correction is necessary. It will be seen from the table that the

Table 5. (*Distribution of N balls in a specified N out of $2N$ compartments.*)
Comparison of exact values and normal approximation for $N = 10$

z	Exact	Normal	z	Exact	Normal
10	0·0010	0·0022	4	0·0901	0·0920
9	0·0156	0·0186	3	0·0176	0·0195
8	0·0881	0·0892	2	0·0016	0·0022
7	0·2334	0·2289	1	0·0001	0·0001
6	0·3194	0·3149	0	0·0000	0·0000
5	0·2332	0·2325			

agreement between exact values and the normal theory approximation is good enough for practical purposes. It is suggested, therefore, that for samples greater than 10 the normal approximation may be used for tests of significance.

15. The derivation of the power function of this second test presents little difficulties in reasoning, although it is laborious to write down. In most combinatorial problems it is often easier to write down the power function and from it deduce the probability distribution of the criterion when the hypothesis tested is true. The fact that the probabilities are different for all compartments makes it difficult to overlook certain combinations which are all too easily passed over when the probability is the same for all compartments. Let the probabilities for the N compartments specified for test be $p_{aT}, p_{bT}, p_{cT}, \ldots, p_{NT}$, and let the probabilities for the other compartments left be $p_{aL}, p_{bL}, \ldots, p_{NL}$. Given these probabilities we have

$$P\{z=N\} = \sum_{a} p_{aL}^{N} + \sum_{r,s}\sum_{a,b} p_{aL}^{r} p_{bL}^{s} + \sum_{r,s,t}\sum_{a,b,c} p_{aL}^{r} p_{bL}^{s} p_{cL}^{t} + \ldots,$$

where $\sum\limits_{a,b,c}$ stands for $\sum\limits_{a=1}^{N}\sum\limits_{\substack{b=1 \\ a\neq b\neq c}}^{N}\sum\limits_{c=1}^{N}$, etc.,

and all summations unless explicitly stated otherwise run from 1 to N. The outside summation $\sum$ in each case is taken over all possible partitions of N, excluding zero as a part. It is easy to show that the above expression reduces to

$$P\{z=N\} = \left(\sum_{a=1}^{N} p_{aL}\right)^{N}.$$

Using the partition notation of § 5, we may write down the other probabilities for z. We give the expressions for $z = N-1$ and $z = N-2$, the others may easily be deduced in a similar fashion:

$$\begin{aligned} P\{z=N-1\} = {} & \sum_{a} p_{aT}^{N} + \sum_{a_1,a_2}\sum_{a}\sum_{b} P(a_1,a_2)\,(p_{aL}^{a_1}p_{bT}^{a_2} + p_{aT}^{a_1}p_{bL}^{a_2}) \\ & + \sum_{a_1,a_2,a_3} P(a_1,a_2,a_3)\left(\sum_{a,b}\sum_{c} p_{aL}^{a_1}p_{bL}^{a_2}p_{cT}^{a_3} + \sum_{a,c}\sum_{b} p_{aL}^{a_1}p_{bT}^{a_2}p_{cL}^{a_3} + \sum_{b,c}\sum_{a} p_{aT}^{a_1}p_{bL}^{a_2}p_{cL}^{a_3}\right) \\ & + \sum_{a_1,a_2,a_3,a_4} P(a_1,a_2,a_3,a_4)\left(\sum_{b,c,d}\sum_{a} p_{aT}^{a_1}p_{bL}^{a_2}p_{cL}^{a_3}p_{dL}^{a_4} + \sum_{a,c,d}\sum_{b} p_{aL}^{a_1}p_{bT}^{a_2}p_{cL}^{a_3}p_{dL}^{a_4}\right. \\ & \qquad \left. + \sum_{a,b,d}\sum_{c} p_{aL}^{a_1}p_{bL}^{a_2}p_{cT}^{a_3}p_{dL}^{a_4} + \sum_{a,b,c}\sum_{d} p_{aL}^{a_1}p_{bL}^{a_2}p_{cL}^{a_3}p_{dT}^{a_4}\right) + \text{etc.} \end{aligned}$$

For the sake of simplifying the arithmetic, the multiple sums can be expressed as products of the one-part symmetric functions (i.e. power sums) from David & Kendall's tables. Again

$$\begin{aligned} P\{z=N-2\} = {} & \sum_{a_1,a_2} P(a_1,a_2) \sum_{a,b} p_{aT}^{a_1}p_{bT}^{a_2} + \sum_{a_1,a_2,a_3} P(a_1,a_2,a_3) \\ & \times\left(\sum_{a,b}\sum_{c} p_{aT}^{a_1}p_{bT}^{a_2}p_{cL}^{a_3} + \sum_{a,c}\sum_{b} p_{aT}^{a_1}p_{bL}^{a_2}p_{cT}^{a_3} + \sum_{b,c}\sum_{a} p_{aL}^{a_1}p_{bT}^{a_2}p_{cT}^{a_3}\right) \\ & + \sum_{a_1,a_2,a_3,a_4} P(a_1,a_2,a_3,a_4)\left(\sum_{a,b} p_{aT}^{a_1}p_{bT}^{a_2}\sum_{c,d} p_{cL}^{a_3}p_{dL}^{a_4} + \text{all possible pairs of products}\right) + \text{etc.} \end{aligned}$$

The final term will be

$$P\{z=0\} = \sum_{a,b,\ldots,N} p_{aT}p_{bT}p_{cT}\cdots p_{NT}.$$

16. The formal expressions for the first two moments of z are immediate. We have, in general, for N balls and k compartments

$$\mathscr{E}(z) = \sum_{a=1}^{k}(1-p_{aT})^N,$$

$$\mathscr{E}(z(z-1)) = \sum_{b=1}^{k}\sum_{\substack{a=1\\ a\neq b}}^{k}(1-p_{aT}-p_{bT})^N.$$

We note that if all the p's are equal to $1/(2k)$, we obtain the moments of § 13. A technique similar to that used previously will give these moments in reasonable form for the case when the difference between the two hypotheses, H_0 and H_1, is small. Thus we write

$$p_i = \frac{1}{2k}+\alpha_i,$$

whence
$$\sum_{i=1}^{k}\alpha_i = \sum_{i=1}^{k}p_i-\tfrac{1}{2},$$

and
$$\sum_{i=1}^{k}\alpha_i^2 = \sum_{i=1}^{k}p_i^2-\frac{1}{k}\sum_{i=1}^{k}p_i+\frac{1}{4k}.$$

Hence, assuming $\Sigma\,\alpha_i^s$ is negligible for $s>2$,

$$\mathscr{E}(z) = \sum_{1}^{k}(1-p_i)^N = \sum_{1}^{k}\left(1-\frac{1}{2k}-\alpha_i\right)^N$$
$$= k\left(1-\frac{1}{2k}\right)^N - N\left(1-\frac{1}{2k}\right)^{N-1}\sum_{i=1}^{k}\alpha_i+\frac{N(N-1)}{2}\left(1-\frac{1}{2k}\right)^{N-2}\sum_{i=1}^{k}\alpha_i^2,$$

where we write i for simplicity instead of the subscript aT. The summation is taken over all the compartments assigned for test. Again

$$\mathscr{E}(z(z-1)) = k(k-1)\left(1-\frac{1}{k}\right)^N - N\left(1-\frac{1}{k}\right)^{N-1}\sum_{\substack{i=1\\ i\neq j}}^{k}\sum_{j=1}^{k}(\alpha_i+\alpha_j)+\frac{N(N-1)}{2}\left(1-\frac{1}{k}\right)^{N-2}\sum_{i\neq j}\sum(\alpha_i+\alpha_j)^2,$$

and
$$\sum_{i\neq j}\sum(\alpha_i+\alpha_j) = 2(k-1)\left(\sum_{i=1}^{k}p_i-\tfrac{1}{2}\right),$$

$$\sum_{i\neq j}\sum(\alpha_i+\alpha_j)^2 = 2(k-2)\left(\Sigma\,p_i^2-\frac{1}{k}\Sigma\,p_i+\frac{1}{4k}\right)+2(\Sigma\,p_i-\tfrac{1}{2})^2.$$

We need thus only evaluate the corrective terms containing the α's in order to obtain the mean and by deduction, the variance, of the distribution of z under H_1. Since we have shown that z is approximately normally distributed under H_0 when $N = k = 10$, and since we have assumed H_1 is not very different from H_0, it appears reasonable to assume that z is approximately normally distributed under H_1 with mean and variance as given above, where k is put equal to N. Such assumptions will enable us to obtain the power of the test for samples of 10 and over without recourse to the laborious working out of the appropriate terms of the power function. When H_1 is very different from H_0 and the power is large—the case most often needed in practice—p_{aT} will usually be very small and expansions therefore of $\mathscr{E}(z)$ and $\mathscr{E}(z(z-1))$ in terms of the p's will not be laborious to evaluate numerically. We note, however, as for the previous test, that it is not known whether the assumption of normality will hold in this case.

17. For the sake of example we compare the power of the zero test to detect an increase in the population mean with that of the power of the test which would undoubtedly be used in practice based on the sample mean. We take the sample size of 5. The power function

obtained by assuming the sample mean is normally distributed about the population mean with standard deviation equal to $\sigma/\sqrt{n}$ is drawn out in Fig. 1, and also $P\{z = 5\}$, both for the alternative hypotheses that the population mean has increased. The risk of rejection of the hypothesis when true has been fixed in both cases at 0·03125. When the orthodox test gives a power of 90 %, the zeroes test has a power of 65 %, a satisfactory result considering the broad class of alternative hypotheses to which the latter test can be made to apply.

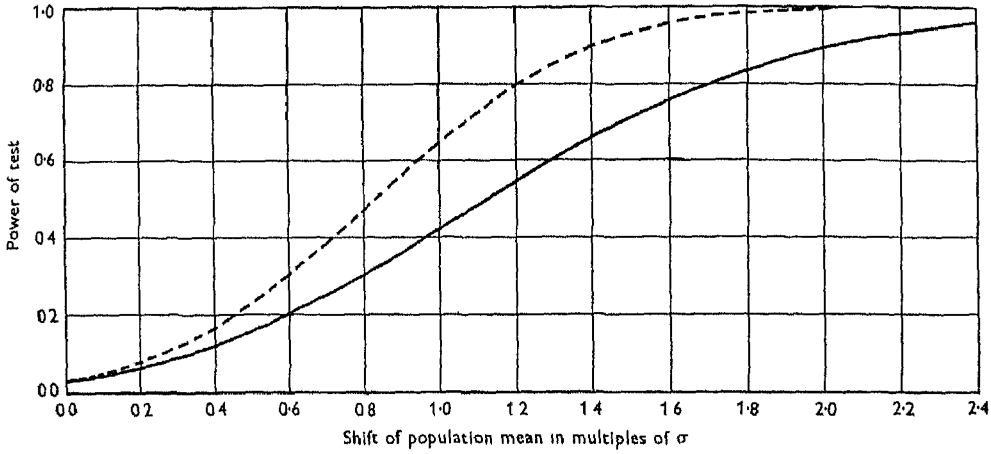

Fig. 1. Power of tests to detect a change in the population mean.
---- orthodox test with sample mean; —— zeroes test.

Prof. Pearson has pointed out to me that in applying the zeroes test in this way, i.e. to detect a shift in the population mean, I am using here an interesting variant of another test. It is known that a quick approximate method of testing whether there has been a shift in the population mean is to see how the sample units fall about the mean, ξ, specified by the hypothesis tested. If all the sample units fall to the right of ξ then the orthodox test using the sample mean will be likely to give a significant result, unless of course there has been at the same time a reduction in the population standard deviation. The zeroes test consists in noticing how the sample units fall about the population median. Table 4*a* shows that for samples of 7 and less it is necessary for all observations to lie on the same side of the median in order to obtain a significant result. For samples of 8, 9 and 10 one observation may lie on the opposite side (or two or three or more provided they are all in the same compartment). For more complex alternative hypotheses the rules that could be laid down become more complex and it is easier to note the groups into which the sample points fall.

I should like to thank Prof. E. S. Pearson for the idea of the second test described here and for much helpful criticism.

REFERENCES

Carver, H. C. (1940). *Statistical Tables*. University of Michigan.
David, F. N. (1949). *Probability Theory for Statistical Methods*. Cambridge University Press.
David, F. N. & Johnson, N. L. (1948). *Biometrika*, **35**, 182.
David, F. N. & Kendall, M. G. (1949). *Biometrika*, **36**, 431.
Elderton, E. M. & Pearson, K. (1925). *Biometrika*, **17**, 200.
Fix, E. (1949). *University of California Publications in Statistics*, **1**, 15.
Patnaik, P. B. (1949). *Biometrika*, **36**, 202.
Pearson, E. S. (1938). *Biometrika*, **30**, 134.
Stevens, W. L. (1937). *Ann. Eugen., Lond.*, **8**, 57.

Table 1*a*. *Probability distribution of z, the number of zero groups, when N sample units are randomly arranged in N groups of equal probability*

N z	3	4	5	6	7	8	9	10	11
0	0·2222	0·0937	0·0384	0·0154	0·0061	0·0024	0·0009	0·0004	0·0001
1	0·6667	0·5625	0·3840	0·2315	0·1285	0·0673	0·0337	0·0163	0·0077
2	0·1111	0·3281	0·4800	0·5015	0·4284	0·3196	0·2164	0·1361	0·0808
3	—	0·0156	0·0960	0·2315	0·3570	0·4206	0·4131	0·3556	0·2770
4	—	—	0·0016	0·0199	0·0768	0·1703	0·2713	0·3451	0·3730
5	—	—	—	0·0001	0·0032	0·0193	0·0606	0 1286	0·2093
6	—	—	—	—	0·0000	0·0004	0·0039	0·0172	0·0479
7	—	—	—	—	—	0·0000	0·0000	0·0007	0 0040
8	—	—	—	—	—	—	0·0000	0·0000	0·0001
9	—	—	—	—	—	—	—	0·0000	0·0000
10	—	—	—	—	—	—	—	—	0·0000

N z	12	13	14	15	16	17	18	19	20
0	0·0001	0·0000	0·0000	0·0000	0·0000	0·0000	0·0000	0·0000	0·0000
1	0·0035	0·0016	0·0007	0·0003	0·0001	0·0000	0·0000	0·0000	0·0000
2	0·0458	0·0250	0·0132	0·0068	0·0033	0·0017	0·0008	0·0000	0·0000
3	0·1994	0·1348	0·0864	0·0530	0·0313	0·0179	0·0100	0·0052	0·0029
4	0·3560	0·3080	0·2461	0·1841	0·1303	0·0880	0·0570	0·0357	0·0216
5	0·2809	0·3255	0·3357	0·3151	0·2735	0·2222	0·1707	0·1248	0·0874
6	0·0988	0·1632	0·2279	0·2784	0·3052	0·3058	0·2839	0·2470	0·2031
7	0·0147	0·0380	0·0768	0·1284	0·1847	0·2353	0·2709	0·2863	0·2811
8	0·0008	0·0038	0·0123	0·0303	0·0602	0·1016	0·1498	0·1981	0·2365
9	0·0000	0·0001	0·0009	0·0035	0·0103	0·0242	0·0476	0·0809	0·1215
10	0·0000	0·0000	0·0000	0·0002	0·0009	0·0030	0·0085	0·0194	0·0378
11	0·0000	0·0000	0·0000	0·0000	0·0000	0·0002	0·0008	0·0026	0·0070
12	—	0·0000	0·0000	0·0000	0·0000	0·0000	0·0000	0·0002	0·0007
13	—	—	0·0000	0·0000	0·0000	0·0000	0·0000	0·0000	0·0000
14	—	—	—	0·0000	0·0000	0·0000	0·0000	0·0000	0·0000
15	—	—	—	—	0·0000	0·0000	0·0000	0·0000	0·0000
16	—	—	—	—	—	0·0000	0·0000	0·0000	0·0000
17	—	—	—	—	—	—	0·0000	0·0000	0·0000
18	—	—	—	—	—	—	—	0·0000	0·0000
19	—	—	—	—	—	—	—	—	0·0000

Table 1*b*. *Suggested significance levels for the distribution of z as given in Table* 1*a*

N	5	6	7	8	9	10
Reject if z is	0 (0·038) 4 (0·002)	0 (0·015) ⩾4 (0·020)	0 (0·006) ⩾5 (0·003)	0 (0·002) ⩾5 (0·020)	⩽1 (0·035) ⩾5 (0·064)	⩽1 (0·017) ⩾6 (0·018)
N	11	12	13	14	15	16
Reject if z is	⩽1 (0·008) ⩾6 (0·052)	⩽2 (0·049) ⩾7 (0·015)	⩽2 (0·027) ⩾7 (0·042)	⩽2 (0·014) ⩾8 (0·013)	⩽3 (0·060) ⩾8 (0·034)	⩽3 (0·035) ⩾9 (0·011)
N	17	18	19	20		
Reject if z is	⩽3 (0·019) ⩾9 (0·027)	⩽3 (0·011) ⩾9 (0·057)	⩽4 (0·041) ⩾10 (0·022)	⩽4 (0·025) ⩾10 (0·045)		

N.B. The figures in parentheses are the resulting probabilities of wrong rejection.

Table 4*a*. *Distribution of z, the number of zero groups, in any specified N out of a total of* 2*N compartments*

z \ N	1	2	3	4	5	6	7	8	9	10
0	0·5000	0·1250	0·0278	0·0059	0·0012	0·0002	0·0001	0·0000	0·0000	0·0000
1	0·5000	0·6250	0·3333	0·1289	0·0420	0·0123	0·0033	0·0008	0·0002	0·0001
2	—	0·0250	0·5139	0·4424	0·2550	0·1163	0·0455	0·0160	0·0052	0·0016
3	—	—	0·1250	0·3604	0·4380	0·3436	0·2058	0·1026	0·0447	0·0176
4	—	—	—	0·0625	0·2326	0·3692	0·3715	0·2793	0·1712	0·0901
5	—	—	—	—	0·0312	0·1427	0·2813	0·3484	0·3179	0·2332
6	—	—	—	—	—	0·0156	0·0846	0·2000	0·2958	0·3194
7	—	—	—	—	—	—	0·0078	0·0489	0·1353	0·2334
8	—	—	—	—	—	—	—	0·0039	0·0278	0·0881
9	—	—	—	—	—	—	—	—	0·0020	0·0156
10	—	—	—	—	—	—	—	—	—	0·0010

Table 4*b*. *Suggested significance levels from Table* 4*a*

Sample size	3	4	5	6	7	8	9	10
Reject if number of zeroes is	0	0	⩽1	⩽1	⩽2	⩽2	⩽3	⩽3
Probability of wrong rejection	0·0278	0·0059	0·0432	0·0125	0·0489	0·0168	0·0501	0·0193
Reject if number of zeroes is	—	4	5	6	7	⩾7	⩾8	⩾9
Probability of wrong rejection	—	0·0625	0·0312	0·0156	0·0078	0·0528	0·0298	0·0165

MAXIMUM LIKELIHOOD AND THE EFFICIENCY OF THE METHOD OF MOMENTS

By L. R. SHENTON

1. Introduction

One of the practical difficulties in estimation by maximum likelihood arises from the fact that, except in special cases, the resulting equations are complicated and not easily solved. The present paper approaches this problem by deriving an expansion based on moment approximations to the likelihood equations. These are also not easy to solve explicitly, but they furnish a method of deriving the efficiency of the method of moments in terms of an expansion of the determinant of the information matrix. The efficiency of moment fitting as applied to Pearson's distributions has been discussed by Fisher (1921), who more recently treated the case of the negative binomial (Fisher, 1941).

2. The likelihood equations and orthogonal polynomials

Let $P(x; \theta_1, \theta_2, \ldots, \theta_s)$ be the probability of the variate x, depending on s parameters θ_j ($j = 1, 2, \ldots, s$). We shall assume that $P(x; \theta_j)$ possesses first derivatives with respect to the θ_j, and that its moments μ'_r and their first derivatives exist and are finite. In this case (Kacmarz & Steinhaus, 1935) it is possible to construct a system of polynomials

$$q_r(x) = \sum_{j=0}^{r} A_{rj} x^j \quad (A_{rr} \neq 0,\ r = 0, 1, 2, \ldots), \tag{1}$$

which are orthogonal with respect to $P(x; \theta)$, satisfying

$$\left.\begin{aligned} \int q_r q_s P(x; \theta)\, dx &= 0 \quad (r \neq s), \\ &\neq 0 \quad (r = s), \end{aligned}\right\} \tag{2}$$

or

$$\left.\begin{aligned} \Sigma q_r q_s P(x; \theta) &= 0 \quad (r \neq s), \\ &\neq 0 \quad (r = s), \end{aligned}\right\} \tag{3}$$

where the integral includes finite summation for a discrete variate. These may be replaced by

$$\begin{aligned} \int q_r \phi_s(x) P(x; \theta)\, dx &= 0 \quad (r > s), \\ &\neq 0 \quad (r = s), \end{aligned}$$

with a similar expression for (3), where $\phi_j(x)$ is an arbitrary polynomial of degree j.

It is convenient to assume for q_r the form

$$q_r(x) = \sum_{j=1}^{r} B_{rj}(x^j - \mu'_j) \quad (r = 1, 2, \ldots;\ q_0 = 1), \tag{4}$$

where μ'_s is the sth crude moment of $P(x; \theta)$.

Choosing $\phi_j(x) = x^j$, the orthogonality conditions lead to

$$\sum_{j=1}^{r} B_{rj}(\mu'_{j+k} - \mu'_j \mu'_k) = 0 \quad (k = 1, 2, \ldots, r-1). \tag{5}$$

$$-\int q_r^2 P(x; \theta)\, dx + \sum_{j=1}^{r} B_{rj}(\mu'_{j+r} - \mu'_j \mu'_r) = 0. \tag{6}$$

Writing m'_s for sample moments, we have

$$N \operatorname{cov}(m'_r, m'_s) = \mu'_{r+s} - \mu'_r \mu'_s = (r, s) \quad \text{say,}$$

where N is the number in the sample, so that from (4) and (5)

$$q_r(x) = \begin{vmatrix} x^r - \mu'_r & x^{r-1} - \mu'_{r-1} & \dots & x - \mu'_1 \\ (r,1) & (r-1,1) & \dots & (1,1) \\ (r,2) & (r-1,2) & \dots & (1,2) \\ \vdots & & & \\ (r,r-1) & (r-1,r-1) & \dots & (1,r-1) \end{vmatrix} \div \begin{vmatrix} (r-1,1) & \dots & (1,1) \\ (r-1,2) & \dots & (1,2) \\ \vdots & & \\ (r-1,r-1) & \dots & (1,r-1) \end{vmatrix}, \tag{7}$$

and from (5) and (6)

$$\int q_r^2 P(x;\theta)\,dx = (-)^{r-1} \begin{vmatrix} (r,1) & (r-1,1) & \dots & (1,1) \\ (r,2) & (r-1,2) & \dots & (1,2) \\ \vdots & & & \\ (r,r) & (r-1,r) & \dots & (1,r) \end{vmatrix} - \begin{vmatrix} (r-1,1) & \dots & (1,1) \\ \vdots & & \\ (r-1,r-1) & \dots & (1,r-1) \end{vmatrix} \tag{8}$$

$$= \phi_r \quad \text{say.}$$

Suppose further that $$\frac{\partial P_x}{\partial \theta_j} = \sum_{k=0} C_{jk} q_k P_x \quad (j = 1, 2, \dots, s), \tag{9}$$

where $P_x \equiv P(x; \theta)$. Then

$$\int \frac{\partial P_x}{\partial \theta_j} q_r\,dx = \phi_r C_{jr} \quad (j = 1, 2, \dots, s;\ r = 0, 1, \dots) \tag{10}$$

$$= \int \frac{\partial P_x}{\partial \theta_j} \sum_{k=1}^{r} B_{rk}(x^k - \mu'_k)\,dx$$

$$= \sum_{k=1}^{r} B_{rk} \frac{\partial}{\partial \theta_j} \int x^k P_x\,dx - \sum_{k=1}^{r} \mu'_k B_{rk} \frac{\partial}{\partial \theta_j} \int P_x\,dx \tag{11}$$

$$= \begin{vmatrix} \dfrac{\partial \mu'_r}{\partial \theta_j} & \dfrac{\partial \mu'_{r-1}}{\partial \theta_j} & \dots & \dfrac{\partial \mu'_1}{\partial \theta_j} \\ (r,1) & (r-1,1) & \dots & (1,1) \\ (r,2) & (r-1,2) & \dots & (1,2) \\ \vdots & & & \\ (r,r-1) & (r-1,r-1) & \dots & (1,r-1) \end{vmatrix} \div \begin{vmatrix} (r-1,1) & \dots & (1,1) \\ (r-1,2) & \dots & (1,2) \\ \vdots & & \\ (r-1,r-1) & \dots & (1,r-1) \end{vmatrix}, \tag{12}$$

since $\dfrac{\partial}{\partial \theta_j} \displaystyle\int P_x\,dx = 0$ provided the range of values of x is independent of the θ_j, and where we have assumed it is permissible to differentiate under the integration sign. Clearly $C_{j0} = 0$. Now the maximum likelihood equations for the estimation of the parameters θ_j are

$$\Sigma' \frac{\partial}{\partial \theta_j} \log P_x = 0 \quad \text{or} \quad \Sigma' \frac{1}{P_x} \frac{\partial P_x}{\partial \theta_j} = 0 \quad (j = 1, 2, \dots, s),$$

the summation being over the sample.

From (9) we have the following formal expansion to determine optimum statistics:

$$C_{j1}(m'_1 - \mu'_1) + C_{j2} \frac{\begin{vmatrix} m'_2 - \mu'_2 & m'_1 - \mu'_1 \\ (2,1) & (1,1) \end{vmatrix}}{(1,1)} + C_{j3} \frac{\begin{vmatrix} m'_3 - \mu'_3 & m'_2 - \mu'_2 & m'_1 - \mu'_1 \\ (3,1) & (2,1) & (1,1) \\ (3,2) & (2,2) & (1,2) \end{vmatrix}}{\begin{vmatrix} (2,1) & (1,1) \\ (2,2) & (1,2) \end{vmatrix}} + \dots = 0$$

$$(j = 1, 2, \dots, s), \tag{13}$$

where C_{jr} is given by (12), and $(k, l) \equiv N \operatorname{cov}(m'_k, m'_l)$. The equations (13) will in general be very difficult to solve, and from a practical point of view of doubtful value. To use them at all we must curtail the series at a certain stage and assume that the remainders are negligible. Further, since the sampling variance of m'_s is large for $s > 4$ or 5, the inclusion of additional terms to improve mathematical accuracy will partly defeat its own object. The most suitable application would appear to be to large samples when the number of parameters is one or two; even in this case the usual method of using an inefficient estimator and successively approximating from the likelihood equation may be simpler. We turn to the main object of the paper which is an expansion for the efficiency of the method of moments.

3. EFFICIENCY OF THE METHOD OF MOMENTS

Using the first s sample moments to estimate the s parameters, we have (Fisher, 1941) for the efficiency E_f

$$\frac{1}{E_f} = \left| E\left[\frac{1}{P_x^2}\frac{\partial P_x}{\partial \theta_j}\frac{\partial P_x}{\partial \theta_k}\right]\right| \begin{vmatrix} \operatorname{var}\bar{\theta}_1 & \operatorname{cov}(\bar{\theta}_1, \bar{\theta}_2) & \dots & \operatorname{cov}(\bar{\theta}_1, \bar{\theta}_s) \\ \operatorname{cov}(\bar{\theta}_2, \bar{\theta}_1) & \operatorname{var}(\bar{\theta}_2) & \dots & \operatorname{cov}(\bar{\theta}_2, \bar{\theta}_s) \\ \vdots & \vdots & & \vdots \\ \operatorname{cov}(\bar{\theta}_s, \bar{\theta}_1) & \operatorname{cov}(\bar{\theta}_s, \bar{\theta}_2) & .. & \operatorname{var}(\bar{\theta}_s) \end{vmatrix}$$

$$= \left| E\left[\frac{1}{P_x^2}\frac{\partial P_x}{\partial \theta_j}\frac{\partial P_x}{\partial \theta_k}\right]\right| \left| [\operatorname{cov}(m'_j m'_k)] \right| \left(\frac{\partial(\theta_1, \theta_2, \,.\,.\,, \theta_s)}{\partial(\mu'_1, \mu'_2, \dots, \mu'_s)}\right)^2, \tag{14}$$

where $\bar{\theta}_j$ are consistent estimates of θ_j obtained from the sample moments, $[v_{jk}]$ is a matrix with $j, k = 1, 2, \dots, s$ and $\dfrac{\partial(\theta_1, \theta_2, \,.\,.\,, \theta_s)}{\partial(\mu'_1, \mu'_2, \dots, \mu'_s)}$ is a Jacobian.

But from (9) we have

$$E\left(\frac{1}{P_x^2}\frac{\partial P_x}{\partial \theta_j}\frac{\partial P_x}{\partial \theta_k}\right) = \int \left(\sum_{m=0}^{\infty} C_{jm} q_m\right)\left(\sum_{n=0}^{\infty} C_{kn} q_n\right) P_x \, dx$$

$$= \sum_{\lambda=1}^{\infty} C_{j\lambda} C_{k\lambda} \phi_\lambda \quad (j, k = 1, 2, \dots, s). \tag{15}$$

Hence from (14) and (15)

$$\frac{1}{E_f} = \phi_1 \phi_2 \dots \phi_s \left(\frac{\partial \theta}{\partial \mu'}\right)^2 \left| \left[\sum_{\lambda=1}^{\infty} C_{j\lambda} C_{k\lambda} \phi_\lambda\right] \right|$$

$$= \phi_1 \phi_2 \dots \phi_s \left(\frac{\partial \theta}{\partial \mu'}\right)^2 \left\{ \begin{vmatrix} C_{11} & C_{21} & \dots & C_{s1} \\ C_{12} & C_{22} & \dots & C_{s2} \\ \vdots & \vdots & & \vdots \\ C_{1s} & C_{2s} & \dots & C_{ss} \end{vmatrix} \phi_1 \phi_2 \dots \phi_s + \begin{vmatrix} C_{11} & C_{21} & \dots & C_{s-1,1} & C_{s+1,1} \\ C_{12} & C_{22} & \dots & C_{s-1,2} & C_{s+1,2} \\ \vdots & \vdots & & \vdots & \vdots \\ C_{1s} & C_{2s} & & C_{s-1,2} & C_{s+1,s} \end{vmatrix}^2 \phi_1 \phi_2 \dots \phi_{s-1} \phi_{s+1} \right.$$

$$\left. + \begin{vmatrix} C_{11} & C_{21} & \dots & C_{s+1,1} & C_{s,1} \\ C_{12} & C_{22} & & C_{s+1,2} & C_{s,2} \\ \vdots & \vdots & & \vdots & \vdots \\ C_{1s} & C_{2s} & & C_{s+1,2} & C_{s,s} \end{vmatrix}^2 \phi_1 \phi_2 \dots \phi_{s-2} \phi_{s+1} \phi_s + \dots \right\}, \tag{16}$$

where we have used $\partial\theta/\partial\mu'$ for the Jacobian appearing in (14), and where the determinants are of order $s \times s$ whose terms consist of all distinct combinations of s from $1, 2, 3, \dots$, etc., used in the determinant

$$\begin{vmatrix} C_{p1} & C_{q1} & C_{r1} & \dots & C_{w1} \\ C_{p2} & C_{q2} & C_{r2} & \dots & C_{w2} \\ \vdots & \vdots & \vdots & & \vdots \\ C_{ps} & C_{qs} & C_{rs} & \dots & C_{ws} \end{vmatrix}^2 \phi_p \phi_q \phi_r \dots \phi_w,$$

with p, q, r, etc., as positive integers.

The first term of (16) after elementary operations on rows and columns of the determinant reduces to

$$\phi_1\phi_2\ldots\phi_s\left(\frac{\partial\theta}{\partial\mu'}\right)^2\left(\frac{1}{\phi_1\phi_2\ldots\phi_s}\frac{\partial\mu'}{\partial\theta}\right)^2\phi_1\phi_2\ldots\phi_s = 1.$$

Since $\phi_r = \int q_r^2 P_x dx$ and P_x is essentially positive for the range of values of x for which it is defined, it is clear that ϕ_r is positive. Hence (16) consists of a series of positive terms, the first few of which will provide a lower bound to $1/E_f$ or upper bound to E_f. In general it is quite sufficient to use the first two or three terms.

4. APPLICATION

As an example of the foregoing we take a 'hypergeometric' distribution recently discussed by J. G. Skellam (1948). From it, as Skellam shows, the negative binomial distribution may be found as a limiting case, for which the efficiency of the method of moments has been given. Taking the binomial distribution

$$\phi(x) = \binom{n}{x}p^x q^{n-x},$$

where the probability varies according to the law

$$\phi(p) = kp^{\alpha-1}(1-p)^{\beta-1} \quad (0 \leqslant p \leqslant 1),$$

with k a constant and α and β any real positive values, Skellam arrives at the probability function

$$P(x;\alpha,\beta) = \binom{n}{x}(\alpha+x-1)^{(x)}\frac{(\beta+n-x-1)^{(n-x)}}{(\alpha+\beta+n-1)^{(n)}} \tag{17}$$

with $$\binom{n}{x} \equiv \frac{n(n-1)\ldots(n-x+1)}{x!} \quad \text{and} \quad N^{(x)} \equiv N(N-1)\ldots(N-x+1).$$

It is assumed that n is a known positive integer, so that there are two parameters α and β for estimation. If the value of n is also to be estimated the derivation of E_f will be slightly more complicated. The distribution (17) may be regarded as a negative hypergeometric, in the sense that in the usual form of the hypergeometric probability distribution

$$\psi(x) = \binom{n}{x}(Np)^{(x)}(Nq)^{(n-x)}/N^{(n)}$$

we need only put $Np = -\alpha$, $Nq = -\beta$. The negative binomial follows by putting $\beta = n/c$ and letting $n \to \infty$. Using known properties of the hypergeometric distribution and its associated orthogonal polynomials (Aitken & Gonin, 1935), we have

$$\sum_{x=0}^{\infty} P_x t^x = F(-n,\alpha;\alpha+\beta;1-t), \tag{18}$$

$$q_r(x) = F(n-r+1,-\alpha-r+1;-\alpha-\beta-2r+2;-\Delta_x)x^{(r)}, \tag{19}$$

$$\phi_r = r!\frac{n^{(r)}(\alpha+r-1)^{(r)}(\beta+r-1)^{(r)}(\alpha+\beta+r+n-1)^{(r)}}{(\alpha+\beta+2r-1)^{(2r)}(\alpha+\beta+2r-2)^{(r)}} \quad (r>0), \tag{20}$$

where $$F(a,b;c,d) = 1+\frac{abd}{1!c}+\frac{a(a+1)b(b+1)d^2}{2!c(c+1)}+\ldots$$

and $$\Delta_x u_x = u_{x+1} - u_x.$$

Putting $$\frac{\partial P_x}{\partial \alpha} = \sum_r A_r q_r P_x, \quad \frac{\partial P_x}{\partial \beta} = \sum_r B_r q_r P_a$$

it may be seen that $$A_0 = B_0 = 0,$$

$$A_r \phi_r = \frac{(-)^{r-1} n^{(r)} (\beta + r - 1)^{(r)} (r-1)!}{(\alpha+\beta+2r-2)^{(r)} (\alpha+\beta+r-1)^{(r)}}, \quad B_r \phi_r = \frac{-n^{(r)} (\alpha + r - 1)^{(r)} (r-1)!}{(\alpha+\beta+2r-2)^{(r)} (\alpha+\beta+r-1)^{(r)}}. \tag{21}$$

Using (18) for the factorial moments we find

$$\frac{\partial(\mu_{[1]}, \mu_{[2]})}{\partial(\alpha, \beta)} = \frac{\partial(\mu_1', \mu_2')}{\partial(\alpha, \beta)} = \frac{-n \,.\, n^{(2)} \alpha\beta}{(\alpha+\beta)^3 (\alpha+\beta+1)^2} = D \quad \text{say},$$

so that estimating α and β by the first two factorial moments of the sample, we have from (20), (21) and (16)

$$\begin{aligned} \frac{1}{E_f} &= 1 + \frac{\phi_1 \phi_2}{D^2} \left[\begin{vmatrix} A_1 & A_3 \\ B_1 & B_3 \end{vmatrix}^2 \phi_1 \phi_3 + \begin{vmatrix} A_2 & A_3 \\ B_2 & B_3 \end{vmatrix}^2 \phi_2 \phi_3 + \begin{vmatrix} A_1 & A_4 \\ B_1 & B_4 \end{vmatrix}^2 \phi_1 \phi_4 + \dots \right] \\ &= 1 + \frac{4(n-2)(\alpha+\beta+1)(\alpha+\beta+3)(\alpha+\beta+5)(\alpha-\beta)^2}{3(\alpha+2)(\beta+2)(\alpha+\beta+2)^4 (n+\alpha+\beta+2)} \\ &\quad + \frac{2(n-1)^2 (\alpha+1)(\beta+1)(\alpha+\beta)(\alpha+\beta+4)^2 (\alpha+\beta+5)}{3(\alpha+2)(\beta+2)(\alpha+\beta+1)^2 (\alpha+\beta+2)^4 (n+\alpha+\beta+2)^{(2)}} \\ &\quad + \frac{3(n-2)^{(2)} (\alpha+\beta+1)(\alpha+\beta+7)[(\alpha+3)^{(3)} + (\beta+3)^{(3)}]^2}{(\alpha+3)^{(2)} (\beta+3)^{(2)} (\alpha+\beta+2)^3 (\alpha+\beta+3)^3 (n+\alpha+\beta+3)^{(2)}} + \dots . \end{aligned} \tag{22}$$

From (22) it can be seen that:

(*a*) The general term associated with $\begin{vmatrix} A_r & A_s \\ B_r & B_s \end{vmatrix}$ $(r \neq s)$ is of the form $K n^{(r)} n^{(s)} / n^2 (n-1)$, where K does not contain n, so that for all admissible suffixes except $r, s = 1, 2$, we see that $(n-2)$ is a factor. Hence when $n = 2$, $E_f = 1$, so that the moment solution is efficient, a fact which is otherwise obvious.

(*b*) If α and β are both small then

$$E_f^{-1} = 1 + \frac{7}{18} \frac{(n-2)(n-3)}{(n+2)(n+3)} + \dots,$$

which for practical values of n, say $\leqslant 20$, suggests that E may not be below 80 %.

(*c*) If α is small, β not small, say $\geqslant 1$, then

$$E_f^{-1} = 1 + \frac{2(n-2)\beta^2 (\beta+1)(\beta+3)(\beta+5)}{3(n+\beta+2)(\beta+2)^5} + \dots,$$

which very rarely reduces E_f below 70 %.

(*d*) If α and β are both large, the relative degrees in α and β of the numerators and denominators suggest high efficiency.

Thus the discrete 'hypergeometric' type distribution (17), depending on two parameters α, β, with range n (assumed known), is simple to fit by the first two factorial moments, and this method of estimating the parameters rarely has low efficiency. (It is suggested that the first four terms of (22) should be used in any practical case.) We may mention that Mood (1943) gives this distribution in connexion with sampling inspection plans, and remarks on some of its properties.

Skellam's example on the secondary association of chromosomes in *Brassica* is

No. of associations $x =$	0	1	2	3
Frequency	32	103	122	80

for which Skellam finds:

(*a*) $\alpha = 6{\cdot}168$, $\beta = 4{\cdot}455$ by moments,

(*b*) 'improved' maximum likelihood estimates, $\hat{\alpha} = 6{\cdot}12$, $\hat{\beta} = 4{\cdot}42$,

(*c*) $\chi^2 = 0{\cdot}76$ for both methods of fitting.

It is clear that the improvement introduced by maximum likelihood methods is negligible, indicating high efficiency for the method of moments. Putting $\alpha = 6$, $\beta = 4{\cdot}5$, $n = 3$ in (22), the first four terms give $E_f = 1/1{\cdot}001$ in good agreement.

It is of interest to notice the form of (22) when we take $\beta = n/c$ in (17) and let $n \to \infty$, so that the distribution is now the negative binomial. We find

$$\frac{1}{E_f} = 1 + \frac{4c}{3(1+c)(\alpha+2)} + \frac{3c^2}{(1+c)^2(\alpha+3)^{(2)}} + \ldots,$$

as given by Fisher (1941) by a different method. Moreover, the likelihood equations (13) take the form

$$\Sigma' \frac{\partial}{\partial c} \log P_x = \Sigma'(x-\alpha c)/c(1+c) = 0,$$

$$\Sigma' \frac{\partial}{\partial \alpha} \log P_x \doteq \Sigma' \left\{ \frac{q_1(x)}{\alpha(1+c)} - \frac{q_2(x)}{2(\alpha+1)^{(2)}(1+c)^2} + \frac{q_3(x)}{3(\alpha+2)^{(3)}(1+c)^3} - \ldots \right\} = 0, \tag{23}$$

where
$$q_r(x) = x^{(r)} - \frac{r}{1!} c(\alpha+r-1)\, x^{(r-1)} + \frac{r(r-1)}{2!} c^2 (\alpha+r-1)^{(2)}\, x^{(r-2)} - \ldots.$$

The first two terms of (23) together with $\Sigma'(x-\alpha c) = 0$ correspond to the moment solution; we may use the first three terms to 'improve' this solution, neglecting the remainder which in any case consists of terms whose expectation is zero. Taking an example given by Haldane (1941):

$x =$	0	1	2	3	4	5	6	7	8	9	Σ
Frequency	167	267	271	185	111	61	27	8	3	1	1096

for which the first three factorial moments are

$$m_{[1]} = 2{\cdot}156{,}934, \quad m_{[2]} = 5{\cdot}100{,}365, \quad m_{[3]} = 12{\cdot}651{,}460,$$

we find from the first three terms of (23)

$$1{\cdot}344{,}003\alpha^3 - 7{\cdot}805{,}548\alpha^2 - 40{\cdot}625{,}167\alpha - 140{\cdot}487{,}797 = 0,$$

a solution of which is $\alpha = 9{\cdot}918$; Haldane gives $\alpha = 9{\cdot}900$ as the maximum likelihood solution and $\alpha = 10{\cdot}44$ as the moment solution, so that the use of (23) does in this case lead to an improvement. It would appear desirable to know under what conditions series such as (23) converge rapidly, and whether it is not just those cases for which the efficiency of moment fitting is low which give slow convergence.

In conclusion I wish to express my thanks to Prof. M. S. Bartlett for helpful criticisms.

REFERENCES*

Aitken, A. C. & Gonin, H. T. (1935). *Proc. Roy. Soc. Edin.* **55**, 114.

Fisher, R. A. (1921). *Philos. Trans.* A, **222**, 309.

Fisher, R. A. (1941). *Ann. Eugen., Lond.*, **11**, 182.

Haldane, J. B. S. (1941). *Ann. Eugen., Lond.*, **11**, 179.

Kacmarz, S. & Steinhaus, H. (1935). *Theorie der orthogonalreihen.* Warsaw.

Mood, A. M. (1943). *Ann. Math. Statist.* **14**, 415.

Skellam, J. G. (1948). *J.R. Statist. Soc.* Series B, **10**, 257.

---

* The method of the present paper is applied to the evaluation of a class of definite integral in a paper to appear in a forthcoming issue of the *Proceedings of the Royal Society of Edinburgh*.

FIDUCIAL LIMITS OF THE PARAMETER OF A DISCONTINUOUS DISTRIBUTION

By W. L. STEVENS*

1. The problem

1·1. It has long been realized that methods of interval estimation which can be used successfully in the case of continuous distributions encounter a special kind of difficulty when we attempt to apply them to a discontinuous distribution such as the binomial or the Poisson. This difficulty has been expressed in different ways by different writers, their choice of words being mainly influenced by whether they are using Fisher's concept of fiducial probability or Neyman's of confidence intervals. While these approaches are said to be fundamentally different, their applications to the uniparametric case are so closely linked that an argument in terms of the one can usually be translated in terms of the other. We shall accordingly not hesitate to express our results in the languages of both theories.

We may begin with the simplest example of a discrete distribution—the binomial. If the probability of an event is π, the probability that it will occur f times in n trials is

$$g(f, \pi) = \frac{n!}{(n-f)!\,f!}(1-\pi)^{n-f}\pi^f. \tag{1·11}$$

The probability that it will happen f *or more* times is

$$G(f, \pi) = \sum_{r=f}^{n} g(r, \pi). \tag{1·12}$$

If in a given experiment it did in fact happen f times, we may define $\pi_0(f)$, corresponding to any probability level, P_0, as the root of the equation

$$G(f, \pi_0) = P_0. \tag{1·13}$$

When $f = 0$, we define $\pi_0(0) = 0.$

Then it can readily be shown that if $f \neq 0$, and if

$$\pi < \pi_0,$$

then $G(f, \pi) < P_0.$

Thus any hypothesis which puts $\pi < \pi_0$ is rejected at the P_0 level of significance in favour of a higher value of π. Hence π_0 is a lower limit to π and may conveniently be termed the lower P_0 limit of π. At the chosen level of significance, we accordingly write either

$$\pi > \pi_0 \quad \text{or} \quad \pi \geqslant \pi_0.$$

Similarly, we may define $\pi_1(f)$, the upper P_1 limit of π, as the root of the equation

$$G(f+1, \pi_1) = 1 - P_1, \tag{1·14}$$

implying that P_1 = probability that r is f *or fewer*.

When $f = n$, we define $\pi_1(n) = 1$. We then write either $\pi < \pi_1$ or $\pi \leqslant \pi_1$. Combining the two inequalities, we therefore have a statement, either

$$\pi_0 < \pi < \pi_1 \quad \text{or} \quad \pi_0 \leqslant \pi \leqslant \pi_1$$

---

* Universidade de São Paulo, Faculdade de Ciências Económicas, Rua Vila Nova, 268, São Paulo, Brazil.

corresponding to any pair of probability levels, subject to the condition that $P_0+P_1<1$. (P_0 and P_1 are usually but not necessarily equal.)

Notice that in calculating either limit the probability of the actual result, f observed in n, is included. This probability is of course finite, whereas if the observations were drawn from a continuous distribution the probability of the actual result would be an infinitesimal. It is from this difference that all the trouble springs.

1·2. *Charts and tables.* Charts of the limits of π as functions of n and f/n for $P_0 = P_1 = \frac{1}{2}$ % and $2\frac{1}{2}$ % were published by Clopper & Pearson (1934). The present writer noticed that if we tabulate $n\pi$ instead of π and use f and f/n instead of n and f/n, we can greatly facilitate interpolation, thus covering in a condensed table all the region where the normal approximation is unsatisfactory. Our table, which covers the range $f = 0$–15, $f/n = 0$–1 and P_0 or $P_1 = 10$, 2·5 and 0·5 % appeared in the second edition of Fisher & Yates.

1·3. *The two fiducial distributions.* Let us now see what happens when we attempt to define the fiducial distribution of π. Differentiating $G(f,\pi)$ with respect to π, we have

$$h_0(\pi)\,d\pi = \frac{\partial G(f,\pi)}{\partial \pi}\,d\pi = \frac{n!}{(n-f)!\,(f-1)!}(1-\pi)^{n-f}\pi^{f-1}\,d\pi. \tag{1·31}$$

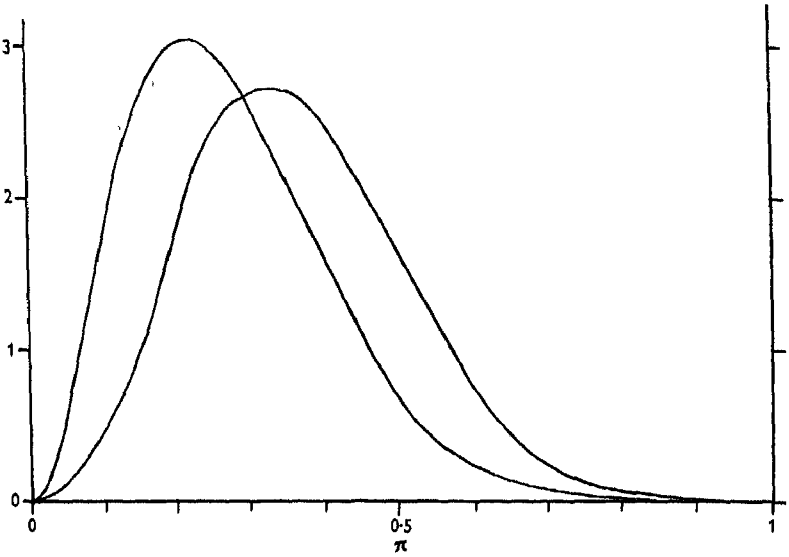

Fig. 1. The pair of fiducial distributions associated with '3 in 10'.

This is an admissible probability distribution, since its integral between 0 and 1 is unity. (See left curve in Fig. 1 for the case $n = 10$ and $f = 3$.) Moreover, if we write

$$\int_0^{\pi_0} h_0(\pi)\,d\pi = P_0, \tag{1·32}$$

the value of π_0 satisfying this equation is in fact the lower P_0 limit of π. If only we could also obtain the upper P_1 limit by solving

$$\int_0^{\pi_1} h_0(\pi)\,d\pi = 1-P_1, \tag{1·33}$$

then we could legitimately call $h_0(\pi)$ the fiducial distribution of π. But, of course, we can do no such thing. We have to define a second distribution of π,

$$h_1(\pi)\,d\pi = \frac{\partial G(f+1,\pi)}{\partial \pi}\,d\pi = \frac{n!}{(n-f-1)!\,f!}(1-\pi)^{n-f-1}\pi^f\,d\pi. \qquad (1\cdot34)$$

(See right curve in Fig. 1.) The upper P_1 limit of π is then the π_1 which satisfies the equation

$$\int_0^{\pi_1} h_1(\pi)\,d\pi = 1-P_1. \qquad (1\cdot35)$$

Thus we see that the unique fiducial distribution obtainable from a continuous distribution (where a sufficient statistic exists) becomes resolved in the discrete case into a *pair* of distributions.

It is a matter of taste, up to this point in the argument, whether we choose to say that there is no fiducial distribution associated with the result, f in n, or whether we say that each result has associated with it a unique pair of fiducial distributions. At any rate, these two between them perform the same service as is performed by the unique fiducial distribution in the continuous case, namely, that of supplying limits to the parameter, corresponding to any pair of significance levels. The only extra rule we need is the simple one, that when finding lower limits we use $h_0(\pi)$ and when upper limits, $h_1(\pi)$.

1·4. *Fisher's discussion of the problem.* The peculiar nature of the fiducial argument when applied to discontinuous data has been expressed differently by Fisher (1935). He was there discussing a fourfold table, showing twins of convicts classified as convicted or not-convicted and monozygotic or dizygotic. The relevant parameter is the cross-product of the cellular probabilities

$$\psi = \frac{\pi_1\pi_4}{\pi_2\pi_3},$$

but it is evident that all his remarks are applicable to the example of the binomial distribution. The passage must be quoted:

> ...We may thus infer that the observations differ significantly at the 1 % level of significance, from any hypothesis which makes ψ greater than 0·48.... This is not a probability statement about ψ. It is a formally precise statement of the results of applying tests of significance. If, however, the data had been continuous in distribution, on the hypothesis considered, it would have been equivalent to the statement that the fiducial probability that ψ exceeds 0·48 is just one chance in a hundred. With discontinuous data, however, the fiducial argument only leads to the result that this probability does not exceed 0·01. We have a statement of inequality, and not one of equality. It is not obvious in such cases, that, of the two forms of statement possible, the one explicitly framed in terms of probability has any practical advantage. The reason why the fiducial statement loses its precision with discontinuous data is that the frequencies in our table make no distinction between a case in which two dizygotic convicts were only just convicted, perhaps on venial charges, or as first offenders, while the remaining 15 had characters above suspicion, and an equally possible case in which the two convicts were hardened offenders, and some at least of the remaining 15 had barely escaped conviction. If we knew where we stood in the range of possibilities represented by these two examples, and had similar information with respect to the monozygotic twins, the fiducial statements derivable from the data would regain their exactitude.

The explanation here advanced for the loss of precision in the fiducial statement may strike the reader as somewhat metaphysical. Although one can think of a continuous phenomenon underlying the discrete classification into convicted and not-convicted, yet in other examples such a notion would at once ring a false note. For example, if the classification is made according to whether the individual possesses or not a certain gene at a certain locus, there can surely be no underlying phenomenon representing all gradations between possessing and not possessing the gene? Be that as it may, it is curious that the method which we are going to develop can be linked up in a striking way with Fisher's concept of underlying continuity.

1·5. *Confidence intervals for π.* Let us next see how the peculiarity of the discontinuous distribution problem reveals itself when we write in terms of confidence intervals. Let us suppose that we make it a rule always to calculate the lower P_0 limit and then to state that

$$\pi > \pi_0 \quad \text{or} \quad \pi \geqslant \pi_0.$$

In following either rule, we shall sometimes make a true statement and sometimes a false one. We can therefore define

$$\text{Prob}\{\pi > \pi_0\} \quad \text{or} \quad \text{Prob}\{\pi \geqslant \pi_0\}$$

as the limit of

$$\frac{\text{number of times the statement is correct}}{\text{number of applications of the rule}}$$

as the number of applications tends to infinity.

As soon as the confidence interval argument was applied to the binomial distribution, it was realized, and in fact pointed out in Clopper & Pearson, that

$$\text{Prob}\{\pi > \pi_0\} \geqslant 1 - P_0 \quad \text{and} \quad \text{Prob}\{\pi \geqslant \pi_0\} > 1 - P_0. \tag{1·51}$$

This probability is a function of the parameter π. The graph of the complementary function, $\text{Prob}\{\pi \leqslant \pi_0\}$ for the case $n = 10$ and $P_0 = 10\ \%$, is given in Fig. 2.

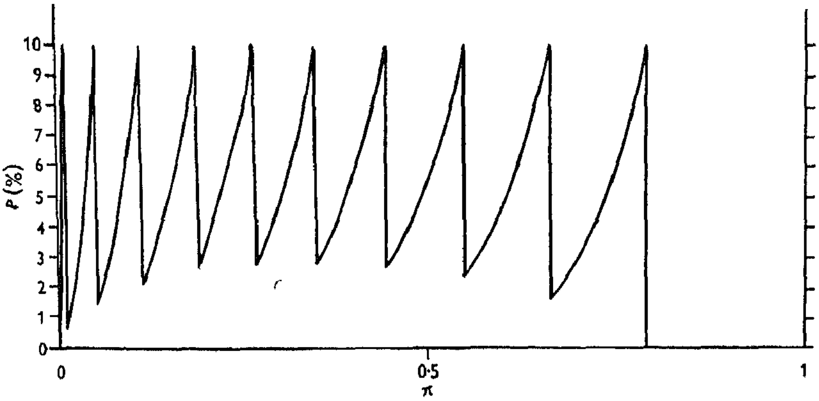

Fig. 2. Probability $\{\pi < \pi_0\}$ when $n = 10$.

The curve shows discontinuities at the n points,

$$\pi_0(1), \quad \pi_0(2), \quad \ldots, \quad \pi_0(n),$$

and it is only at these points that $\text{Prob}\{\pi \leqslant \pi_0\} = P_0$ and hence $\text{Prob}\{\pi > \pi_0\} = 1 - P_0$. Elsewhere, $\text{Prob}\{\pi \leqslant \pi_0\}$ can differ very gravely from 10 %. Leaving out the extreme cases of $f = 0$ and $f = 10$ where we might have expected trouble, we are still left with a variation from $\frac{1}{2}$ to 10 %.

Similarly, it can be shown that, having calculated an upper P_1 limit, π_1, then

$$\text{Prob}\{\pi < \pi_1\} \geqslant 1 - P_1 \quad \text{and} \quad \text{Prob}\{\pi \leqslant \pi_1\} > 1 - P_1. \tag{1·52}$$

The equality sign holds only at the n points of discontinuity,

$$\pi_1(0), \quad \pi_1(2), \quad \ldots, \quad \pi_1(n-1).$$

Combining (1·51) and (1·52), we conclude that

$$\left.\begin{aligned} &\text{Prob}\{\pi_0 < \pi < \pi_1\} \geqslant 1 - P_0 - P_1 \\ \text{and} \quad &\text{Prob}\{\pi_0 \leqslant \pi \leqslant \pi_1\} > 1 - P_0 - P_1. \end{aligned}\right\} \tag{1·53}$$

We can say further that there 'always'* exists a positive quantity D greater than zero, such that

$$\text{Prob}\{\pi_0 < \pi < \pi_1\} > 1 + D - P_0 - P_1. \tag{1·54}$$

Now the use of $1 - P_0 - P_1$ (the *confidence coefficient*) as a measure of our confidence in the assertion, $\pi_0 < \pi < \pi_1$, has been justified by two lines of argument: (*a*) that in the limit, when n is large, the inequality sign may be replaced by the equality sign, and (*b*) that the calculated confidence coefficient is necessarily conservative, i.e. on the 'safe side', because the true probability of π being between the limits is 'always' greater than is stated. The first type of argument need not be taken too seriously, since it amounts to no more than saying that any small-sample test tends in the limit to the corresponding large-sample test, which, although true, would hardly be a reason for ignoring the small sample test if it were available. The second type of justification is a much more serious matter and, in the view of the present writer, is completely invalid.

It is the very basis of any theory of estimation, that the statistician shall be permitted to be wrong a certain proportion of times. Working within that permitted proportion, it is his job to find a pair of limits as narrow as he can possibly make them. If, however, when he presents us with his calculated limits, he says that his probability of being wrong is less than his permitted probability, we can only reply that his limits are unnecessarily wide and that he should narrow them until he is running the stipulated risk. Thus we reach the important, if at first sight paradoxical conclusion, that *it is a statistician's duty to be wrong* the stated proportion of times, and failure to reach this proportion is equivalent to using an inefficient in place of an efficient method of estimation.

Of course, this criticism begs the fundamental question of whether we can, while still not exceeding the stipulated probabilities of being wrong at either limit, find an interval lying within the interval found by the usual method. We shall now show that we can do this and more. In fact, it is possible to find a pair of limits π_0 and π_1 such that

$$\begin{aligned} &\text{Prob}\{\pi \geqslant \pi_0\} = 1 - P_0, \\ &\text{Prob}\{\pi \leqslant \pi_1\} = 1 - P_1, \\ \text{and hence} \quad &\text{Prob}\{\pi_0 \leqslant \pi \leqslant \pi_1\} = 1 - P_0 - P_1, \end{aligned} \tag{1·55}$$

and these limits will 'always' be wholly inside the limits customarily calculated. This means that we can recover for the discontinuous problem the exact type of fiducial or confidence interval statement which is possible in continuous problems.

---

* We shall use inverted commas and write 'always' when we mean that the probability of the event is unity but that it need not happen. Similarly, we shall say an event 'never' happens when it can happen but its probability is zero. Thus if x is drawn at random from the rectangular distribution $0 \leqslant x < 1$, we say that x is 'always' greater than zero, 'never' equal to zero, but never equal to unity.

2. The solution

2·1 Suppose that in n trials the event has occurred f times. Then define a variable y by the equation

$$y = f + x, \tag{2·11}$$

where x is any number chosen at random from the rectangular distribution, $0 \leqslant x < 1$. (This means, in practice, that x is a decimal point followed by a sequence of digits taken from a table of random numbers.) The quantity y is accordingly any real number in the range $0 \leqslant y < n+1$. We may also note that y uniquely determines f and x, since f is the whole and y the fractional part of y.

Since y is the sum of two quantities, each drawn from a known distribution, it follows that its own distribution is determinate, and can in fact be found without difficulty. Now since y uniquely determines f, it is evidently a sufficient statistic for π. We therefore need not hesitate to apply the fiducial probability argument. Thus, if

$$y = f + x$$

and

$$y_0 = f_0 + x_0,$$

$$\begin{aligned} \text{Prob}\{y \geqslant y_0\} &= \text{Prob}\{f > f_0\} + \text{Prob}\{f = f_0\}\,\text{Prob}\{x \geqslant x_0\} \\ &= G(f_0+1, \pi) + g(f_0, \pi)\,(1 - x_0) \\ &= x_0 G(f_0+1, \pi) + (1 - x_0)\,G(f_0, \pi). \end{aligned} \tag{2·12}$$

Differentiating this probability with respect to π, and dropping the suffices on f and x, we thus define a distribution function

$$h_x(\pi)\,d\pi = x h_0(\pi)\,d\pi + (1 - x)\,h_1(\pi)\,d\pi, \tag{2·13}$$

where $h_0(\pi)$ and $h_1(\pi)$ are the functions already defined in (1·31) and (1·34).

We propose to call $h_x(\pi)\,d\pi$ the fiducial distribution of π. Notice, however, that it is defined only after we have drawn our value of x. Given merely the number f, then there exists an infinite continuum of such distributions ranging between the pair of distributions $h_0(\pi)\,d\pi$ and $h_1(\pi)\,d\pi$ which appeared in the earlier discussion. Now it is essential that the fiducial distribution of a parameter should be unique. This, however, can be assured by the following simple rule:

When any experiment has been performed (or series of observations taken), the investigator is allowed once and once only to select, at random, his value of x; the distribution thus determined will be called *the* fiducial distribution of π, and neither he nor anyone else is permitted another drawing of the number x.

As a result of this rule, we find that a fiducial distribution is uniquely determined for each experiment, though not for each value of f. Now it will readily be seen that the fiducial limits found by integrating $h_x(\pi)\,d\pi$ 'always' lie wholly inside the fiducial limits found by the usual method. Moreover, since y is a continuous variable, the statements in fiducial probability become statements of equality instead of statements of inequality.

2·2. *Demonstration.* To clarify these remarks, let us first examine the functional relation between y and π. Corresponding to any probability P (for a lower limit) or $1 - P$ (for an upper limit), we have the following equation relating y and π:

$$xG(f+1, \pi) + (1 - x)\,G(f, \pi) = P. \tag{2·21}$$

Differentiating totally with respect to x, we get

$$\frac{dP}{dx} = \frac{\partial P}{\partial x} + \frac{\partial P}{\partial \pi}\frac{d\pi}{dx} = 0.$$

Hence $$-g(f)+h(\pi)\frac{d\pi}{dx}=0. \qquad (2{\cdot}22)$$

Hence $d\pi/dx$ is everywhere finite and positive.

However, if $y<1-P$, i.e. if $f=0$ and $x<1-P$, the equation becomes

$$(1-\pi)^n=(1-P)/x,$$

thus yielding a negative value for π. We must therefore define π as zero, when $y<1-P$. Similarly, we define π as unity when $y>n+1-P$.

Next we note that as $x\rightarrow 1$ the equation tends to

$$G(f+1,\pi)=P,$$

which is the equation appropriate to $y=f+1$. Hence π is a continuous function of y, although the derivative $d\pi/dy$ is discontinuous at the integral values of y. The graph of π against y is in fact a series of ascending loops, like a telegraph wire going up a mountain side. We conclude therefore that y uniquely determines π and that π, if not equal to zero or unity, also uniquely determines y. Under this same restriction, we also see that $\pi(y)$ and $y(\pi)$ are ever-increasing functions of each other. Moreover, since the lower limit, as usually calculated, can be found from the method here proposed by putting $x=0$, it follows that the new method 'always' yields a higher value for the lower limit. Similarly, it can be shown that it yields a lower value for the upper limit.

Now let π $(0<\pi<1)$ represent the true parametric value, and let $y=y(\pi)$ be the corresponding y at a chosen level P_0. Suppose we draw at random a sample of n observations and encounter the event f_0 times in the sample. To this f_0 we add our randomly chosen x_0 in order to produce y_0, which then uniquely determines π_0, the lower P_0 limit. In repeated applications of this procedure, we may accordingly define

$$\text{Prob}\{\pi_0<\pi\}.$$

However, $\pi_0<\pi$ implies $y_0<y$. Hence

$$\text{Prob}\{\pi_0<\pi\}=\text{Prob}\{y_0<y\}=1-\text{Prob}\{y_0\geqslant y\}=1-P_0. \qquad (2{\cdot}23)$$

Hence Prob $\{\pi_0<\pi\}$, in the sense of the Neyman-Pearson theory, is *exactly* equal to $1-P_0$. Since y is continuous it is immaterial whether we write

$$\text{Prob}\{\pi_0<\pi\}=1-P_0 \quad \text{or} \quad \text{Prob}\{\pi_0\leqslant\pi\}=1-P_0.$$

Similarly, if π_1 is the upper P_1 limit, then

$$\text{Prob}\{\pi<\pi_1\}=\text{Prob}\{\pi\leqslant\pi_1\}=1-P_1. \qquad (2{\cdot}24)$$

Combining the two statements, we therefore have

$$\text{Prob}\{\pi_0<\pi<\pi_1\}=\text{Prob}\{\pi_0\leqslant\pi\leqslant\pi_1\}=1-P_0-P_1. \qquad (2{\cdot}25)$$

These equations hold for all values of π in the open interval $0<\pi<1$. Is it possible to close the interval? Yes, if we adopt the following rules:

When $\pi_0=0$, we must write $0\leqslant\pi$.
When $\pi_0=1$, we must write $1<\pi$.
When $\pi_1=0$, we must write $\pi<0$.
When $\pi_1=1$, we must write $\pi\leqslant 1$.

Under these rules, it can be shown that the equations are true, even for $\pi=0$ and $\pi=1$. However, the solution is a little artificial, since in two instances we have to make statements which we know to be false. Perhaps the proper answer is that in practice we can rule out the possibility of either $\pi = 0$ or $\pi = 1$, for if, in fact, we thought that π was small and possibly zero (or large and possibly unity) we would have employed a different experimental technique, namely that of inverse sampling.

2·3. *Alternative solutions.* Once the idea of introducing a random variate has been admitted, one may inquire whether it need have been drawn from a rectangular distribution. Could it not have been drawn from some other distribution, and if so, could we not find some distribution which would result in yet narrower limits?

Suppose z to be drawn from *any* continuous distribution, $j(z)\,dz$, over *any* range. Then we may consider the joint distribution of the pair of sampling variables, f and z. Let us say, conventionally, that

$$(f, z) > (f_0, z_0)$$

means that either

$$f > f_0$$

or

$$f = f_0 \quad \text{and} \quad z > z_0.$$

From this we have

$$\text{Prob}\{(f, z) > (f_0, z_0)\} = \int_{-\infty}^{z_0} j(z)\,dz\,G(f_0+1, \pi) + \int_{z_0}^{\infty} j(z)\,dz\,G(f_0, \pi). \tag{2·31}$$

Differentiating with respect to π and removing the suffices, we would again have a fiducial distribution of π, which might be offered as an alternative to $h_x(\pi)\,d\pi$.

However, a moment's thought will show that this method yields the same solution in a different guise. In order to draw z_0 at random from $j(z)\,dz$ we would have to select x_0 at random from the rectangular distribution $0 \leqslant x_0 < 1$, and then solve the equation

$$x_0 = \int_{-\infty}^{z_0} j(z)\,dz. \tag{2·32}$$

Replacing the integrals in equation (2·31), we obtain

$$\text{Prob}\{(f, z) > (f_0, z_0)\} = x_0 G(f_0+1, \pi) + (1-x_0)\,G(f_0, \pi). \tag{2·33}$$

This is the same equation as we used before (2·12). Hence we would obtain the same fiducial distributions with the same frequency, the only difference being that they would be differently labelled.

On the other hand, let us consider what happens when the distribution from which we draw z contains points of concentration, e.g. a point $z = Z$, where $\text{Prob}\{z = Z\} = \alpha$, a finite quantity. If then we write

$$x_0 = \int_{-\infty}^{z_0} j(z)\,dz,$$

there will be a one-one relation between x_0 and z_0 everywhere except when $z_0 = Z$, in which point there is a series of values of x_0, $X \leqslant x_0 \leqslant X+\alpha$, corresponding to the one value of z_0. As a result, we find that in this range of x_0, the lower limit remains constant at the value $\pi(f+X)$, i.e. instead of continuing to rise smoothly, the graph of π_0 against x_0 runs horizontally for a space. Within this interval, it is evident that the calculated lower limits will 'always' be below those found from the rectangular distribution. It can also be shown that the inequality sign would have to be returned to the fiducial probability statement if the distribution had points of concentration.

Since we have now shown that alternative methods will either yield the same result, or a result which is in two respects less satisfactory, we conclude that we were justified in calling $h_x(\pi)\,d\pi$ *the* fiducial distribution of π.

2·4. *'Underlying continuity.'* We have seen that the method here proposed replaces integral values of f in the range $0 \leqslant f \leqslant n$, by the continuous variable y in the range $0 \leqslant y < n+1$. We notice at once the link with Fisher's notion (quoted earlier) of the discrete variable being merely an incomplete manifestation of a continuous phenomenon. At the same time, we must emphasize that the method does not imply, nor does it rest on any idea of, underlying continuity. This is proved by the discussion of the preceding section, which showed that the x could be drawn from any continuous distribution. Nevertheless, it is interesting that the method, in its simplest form, is expressible in terms of a continuous variable over the range of f.

One should also notice here a similarity with a method proposed by Fisher (1935) for recovering exact fiducial statements when estimating bacterial density by the dilution method. If the dilution ratio is r, Fisher showed that the fiducial statements may be made exact if the first dilution ratio is taken as r^x instead of r, where x is drawn at random between 0 and 1. There is, however, this essential difference: in Fisher's example a random element is introduced before the experiment is performed, whereas here we propose the introduction of a random element after the experiment is completed.

2·5. *Choice of the x for different limits.* A question which may be asked is whether, when calculating a number of limits, we should select the x once and use it for all limits, or whether we can select a different x for each limit. For the purpose of defining a fiducial distribution, it is obvious that x must be chosen once only. On the other hand, if we wish to determine one or more *confidence intervals*, we may choose, at random, a fresh x for every limit. The intervals so determined will still be genuine confidence intervals in the sense used by Neyman. Nevertheless, it should not be overlooked that such a procedure can lead to paradoxical results. Thus we might find a 90 % confidence interval lying wholly inside an 89 % confidence interval, i.e. we would have to attach more confidence to the more precise statement. Paradoxical though it may sound, the result would be genuine; one would, however, be obliged to make both statements in order to maintain the right proportions of true statements in the long run. Still one would not like to be the statistician who is called upon to explain such a conundrum to a factory manager who has merely been inquiring into the percentage of defective articles in production! Perhaps it would be wise to adopt the rule that the same x is to be used for all limits (in any given experiment), thus bringing the confidence intervals into line with the fiducial intervals.

3. APPLICATION

3·1. There is no particular obstacle, except time, to tabulating the proposed solution, as applied to the problem of the binomial distribution, or indeed to any other discontinuous distribution. Table 1 gives a sample tabulation for the binomial with index = 10. Since the fiducial distribution is now unambiguous, the percentile points are more logically labelled continuously from 0 to 100 %. The upper 10 % limit is of course tabled under 90 %.

Thus suppose 3 events were encountered in 10 trials. A table of random numbers begins: 337... We round this to 34 and enter the table with $f = 3$ and $x = 0·34$, finding the upper and lower 10 % limits

$$13·1 < \pi < 49·4 \quad (\%).$$

For comparison, the usual method gives the wider limits,

$$11·6 < \pi < 55·2 \quad (\%).$$